



U. S. NAVAL SUBMARINE MEDICAL CENTER

Submarine Base, Groton, Conn.

REPORT NUMBER 511

PERSONNEL SELECTION IN THE U. S. NAVY:—

Proceedings of a Navy-Wide Workshop. at the Submarine
Medical Center, April 1967

Benjamin B. Weybrew, Editor

Bureau of Medicine and Surgery, Navy Department
Research Work Unit MF022.01.02-9004.01

Released by:

Gerald J. Duffner, CAPT MC USN
COMMANDING OFFICER
23 February 1968

APPROVED FOR PUBLICATION
RELEASE - DISTRIBUTION
UNLIMITED



**PERSONNEL SELECTION
IN THE U.S. NAVY:
PROCEEDINGS OF A NAVY-WIDE WORKSHOP,
SUBMARINE MEDICAL CENTER, GROTON, CONNECTICUT
APRIL 1967**

Edited by:



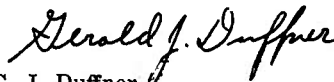
B. B. Weybrew, Ph.D.
Head, Personnel Research Branch
SubMedResLab

Reviewed by:



C. F. Gell, MD, DSc (Med)
Scientific Director
SubMedResLab
Chairman of Workshop

Approved by:



G. J. Duffner
CAPT MC USN
Commanding Officer
Naval Submarine Medical Center
Groton, Connecticut

ADMINISTRATIVE INFORMATION

The Proceedings of the Navy-Wide Workshop on Personnel Selection has been designated as Report No. 1 on Bureau of Medicine and Surgery Research Work Unit MF022.01.02-9004—Selection and Retention of Submarine and Diving Personnel. It was approved for publication on 23 February 1968, as Submarine Medical Research Laboratory Report No. 511. Sixteen previous reports in this area were issued under Work Units MF022.03.03-9021 and MR005.14-2100.1.

PUBLISHED BY THE NAVAL SUBMARINE MEDICAL CENTER

This document has been approved for public release and sale;
its distribution is unlimited.

PREFACE

This workshop was inspired by the Bureau of Medicine and Surgery following a discussion between Captain J. R. Kingston, MC, USN, of ONR, Dr. Charles F. Gell of the Submarine Medical Research Laboratory, Groton, Connecticut, and the signer of this preface. The motivating factor of the original discussions was the question of adequacy existing in the exchange of information between the several Naval Personnel Selection facilities under BuMed administration. A workshop proposal in personnel selection methods in the Navy evolved from these discussions, and the Submarine Medical Center was officially requested to carry out the duties of organizing and conducting such a workshop. The ensuing success of this workshop is an indicator of the most satisfactory approach to information exchange, that is, eyeball-to-eyeball discussion. As a consequence, the Bureau of Medicine and Surgery hopes for an on-going program of planned meetings, workshops and symposia in other fields of endeavor.

The following proceedings of the workshop on personnel selection held at SMRL consists of edited copy of the output from each workshop session taped on the spot. The overall substantive character of these proceedings would seem to attest to the desirability of the workshop mode of information exchange, without the requirement of formally prepared papers and with the goal of conducting principally and primarily a forum for the exchange of ideas.

J. P. POLLARD, CAPT MC USN
Director, Research Division
Bureau of Medicine and Surgery

ACKNOWLEDGMENTS

Appreciation is extended to Dr. Benjamin Weybrew, Head Personnel Research Branch, Submarine Medical Research Laboratory and his staff particularly Dr. Rupert Hester and Mr. James Parker for much of the effort necessary to select the subject matter areas for the program, for participant nominations and contact, and for managing the numerous and often tedious details of running a meeting of this nature. Finally, we are indebted to Mrs. Claire Renkiewicz, Secretary to the Head, Personnel Research Branch for negotiating the numerous letter and telephone communications with the twenty-odd participants necessary in planning this conference and to Miss Shelley Betts for accomplishing the arduous task of transcribing the Workshop tapes.

CHARLES F. GELL, MD, DSc
Chairman

PARTICIPANTS

NAME	NAVAL ACTIVITY REPRESENTED
Miss Rosalie K. Ambler	Naval Aerospace Medical Institute Pensacola, Florida
Dr. Randall M. Chambers	Naval Air Development Center Johnsville, Pennsylvania
CAPT Ralph Christy, MC, USN	Bureau of Medicine and Surgery
LCDR Thomas Gallagher, MSC, USN	Bureau of Medicine and Surgery
Dr. Charles F. Gell	Naval Submarine Medical Research Lab. Groton, Connecticut
LT Leonard R. Green, MSC, USNR	Office of Naval Research
Dr. E. K. Eric Gunderson	Navy Medical Neuropsychiatric Research Unit, San Diego, Calif.
Dr. William W. Haythorn	Naval Medical Research Institute Bethesda, Maryland
Dr. Rupert Hester	Naval Submarine Medical Research Lab. Groton, Connecticut
Dr. William A. Hunt	Northwestern University Evanston, Illinois
LCDR Samuel C. Klagsbrun, MC, USN	Naval Submarine Medical Research Lab. Groton, Connecticut
Dr. H. Barry Molish	Ribicoff Research Center Norwich, Connecticut
Mr. John W. Molyneaux	Chief of Naval Operations
Dr. William Moonan	Naval Personnel Research Activity San Diego, California
LCDR Paul D. Nelson, MSC, USN	Bureau of Medicine and Surgery

CDR Pat O'Connell, MC, USN	Naval Academy, Annapolis, Md.
Mr. James W. Parker	Naval Submarine Medical Research Lab. Groton, Connecticut
Dr. John A. Plag	Navy Medical Neuropsychiatric Research Unit, San Diego, Calif.
Dr. Roland W. Radloff	Naval Medical Research Institute Bethesda, Maryland
Dr. Bernard Rimland	Naval Personnel Research Activity San Diego, California
Dr. Karl E. Schaefer	Naval Submarine Medical Research Lab. Groton, Connecticut
Dr. Gilbert C. Tolhurst	Office of Naval Research
Dr. Benjamin B. Weybrew	Naval Submarine Medical Research Lab. Groton, Connecticut
LCDR Robert J. Wherry, Jr., MSC, USN	Naval Missile Center, Point Magu
Dr. Martin F. Wiskoff	Bureau of Naval Personnel

TABLE OF CONTENTS

	Page
LIST OF PARTICIPANTS	v
INTRODUCTORY REMARKS	1
Welcome and Introductory Comments	1
Captain C. L. Waite MC USN, Commanding Officer	
Naval Submarine Medical Center	
Comments by Workshop Chairman	1
C. F. Gell MD DSc, Chief Scientist	
Submarine Medical Research Laboratory	
KEYNOTE ADDRESS: Some Comments on Military Selection	2
Dr. William A. Hunt, Chairman, Department of	
Psychology, Northwestern University	
SESSION I Brief Updating of Selection Programs by Representatives	
of Participating Activities	11
Some Brief Comments on the Selection Program at the	
Aerospace Medical Institute, Pensacola	11
Rosalie Ambler	
Discussion	
Naval Recruit Selection Program at the Neuropsychiatric	
Research Unit, San Diego	15
Dr. John Plag	
Discussion	
Selection for the Naval Academy	20
CDR Pat O'Connell MC USN	
Selection for Special Assignments: Aquanaut Selection	24
Dr. Roland Radloff, Naval Medical Research	
Institute, Bethesda	
The Submariner Selection Program at New London	29
Dr. Benjamin Weybrew, Naval Submarine Medical Center	
SESSION II Criterion Development	37
Discussion Leader: Dr. Randall Chambers, USN Air	
Development Center, Johnsville	
Overview by Discussion Leader	37
Free Interchange on Session Topic by	
Participants	44

	Page
SESSION III Personality and Aptitude Measurement in Selection	51
Discussion Leader: Dr. Eric Gunderson, Naval Medical Neuropsychiatric Research Unit, San Diego	
Overview by Discussion Leader	51
Free Interchange on Session Topic by Participants	56
SESSION IV Biomedical Indices in Selection	65
Discussion Leader: Dr. Gilbert Tolhurst, Office of Naval Research	
Overview by Discussion Leader	65
Free Interchange on Session Topic by Participants	69
SESSION V Small Groups Observational Data in Selection	77
Discussion Leader: Dr. William Haythorn, Naval Medical Research Institute, Bethesda	
Overview by Discussion Leader	77
Free Interchange on Session Topic by Participants	81
SESSION VI Personnel Allocation Processes	97
Discussion Leader: Dr. William Moonan, Naval Personnel Research Activity, San Diego	
Overview by Discussion Leader	97
Free Interchange on Session Topic by Participants	104
SESSION VII Selection-related Research Areas and Division of Labor among the Navy Activities	111
Discussion Leader: LCDR Paul Nelson, MSC USN Bureau of Medicine and Surgery	
CONCLUDING REMARKS BY WORKSHOP CHAIRMAN	113
APPENDIX A: Results of Pre-workshop Survey of Participants as to the Most Relevant Selection-related Areas for Consideration	117

WELCOME AND INTRODUCTORY COMMENTS

Captain C. L. Waite MC USN, Commanding Officer
US Naval Submarine Medical Center

Ladies and Gentlemen: It is my pleasure to welcome you to the Submarine Medical Center, your host for this Navy-wide Workshop on Personnel Selection. The opportunity to communicate is frequently lost in the large and complex organization which is the modern Navy, and in the daily business of solving our own problems with our limited resources we can seldom afford the luxury of taking time out to look at the overall state of affairs, even in our own specialties. I know that you will not regret putting aside the time for this meeting, for each of you has something to contribute, and I am sure each of you will take something worthwhile home with you.

Upon looking at the program after it was organized, I was greatly impressed with the quality and the variety of talent the Navy has at its disposal. This resulted in the realization of the important role each of you plays, roles encompassing every major operational effort in the Navy. A special note of welcome is due to our guest speaker, Dr. William Hunt, who has taken time out to keynote this meeting. We are indeed honored to have you aboard, Dr. Hunt.

Looking at the selection of Naval personnel historically, at first glance you might say that we have come a long way since the early impressment gangs that scoured the waterfront with no criteria and only a quota of seafaring men to fill. Yet in light of some current methods of manpower procurement within the operating forces, Army, Navy, Air Force for example, the existing selective service program in no small measure suggests that perhaps we haven't progressed too far from the "sound of wind and limb" days. With terrestrial and extraterrestrial conditions being as they are we can no longer afford the "two years before the mast" natural shakedown of separating the men from the boys. Now, men selected and trained at considerable expense must fit into the complex

weapon system of an aircraft, a ship, a submarine or a combat team in almost assembly line fashion. Such is the harsh world with its unreasonable ground rules with which we must deal. We must be realists and practical with respect to selection problems, but to me, in this instance, practicality means not giving ground on major selection criteria.

I will not say anything specific about submarine or diving selection, in which I am keenly interested, since this will be covered by Drs. Weybrew and Radloff later this morning. Obviously retention is closely bound to selection. Perhaps we should talk about matters related to this problem. There is no question at all of the situation being eased if we could retain proportionately more of the people we so carefully select and lavishly train. Possibly Dr. Weybrew's recent work with the Institute of Naval Studies will begin to bear fruit in connection with the complex retention problem. So, let's get on with the business of the day. I sincerely hope you find this workshop as productive as Captain Pollard and I are confident it will be. Again, welcome to New London. Our house is yours.

COMMENTS BY WORKSHOP CHAIRMAN C. F. Gell MD DSc

Captain Waite, Conferees and Guests: Perhaps you are wondering why I, one of the lesser lights of your discipline, stand before you as Chairman of your august body. If you are puzzling on this matter, let me state that I believe Captain Pollard felt I didn't know enough about personnel selection to gum up the meeting, but at the same time I was big enough to step between any belligerents who might face each other to emphasize their discussion. As stated by Captain Waite, this Workshop was recommended by Captain Pollard and Captain Kingston as a means of

exchanging ideas among the personnel engaged in selection of people required for specific jobs in the Navy.

You may ask yourself why we are assembled in a workshop rather than a more formal meeting. The answer to this is simply that a workshop is task oriented and as such we can discuss the problems that beset the various programs. For instance, what were your successes and failures and why did they come about? There will be no formal papers, rather the disgorging of ideas; in fact, you might say idea-sharing is the goal of this workshop. For example, this workshop may give you an opportunity to raise and open for discussion a new item analytical technique, or perhaps a new test or personality concept overlooked in the past. This is an opportunity to resurrect some old data that for a variety of reasons you did not publish. We should make these sessions a "no holds barred" program. We should not suppress any reasonable idea that has as much as a fraction of 1% probability of being useful in one situation or another.

I am sure that most of you will agree that personnel selection as a discipline has a public image problem. There has been criticism particularly in the area of personality testing. The critics have become vociferous and have in some cases rather sullied the public image of the psychological fraternity, particularly in the testing and personality assessment areas. It is quite possible that ideas originating in these discussions may lead to insights, and ideally to research programs that may help to close the credibility gap between test constructors and the users of the test data. I expect that our principal speaker, who was chosen with a definite purpose in mind, will have some words of wisdom and state some original thoughts on these matters.

Now about the workshop program itself; let me speak for a moment on the format for the discussion sessions to follow. You will note on your programs that each session has a discussion leader. The leader was selected on the basis of his background of experience and training to set the tone for the sessions. It will be his job also to channel the discus-

sion so that it remains within the scope of the specific problem area.

And now since we are warmed up sufficiently to be receptive to new ideas drawn from extensive experience in the field of personnel selection, let me introduce our distinguished key note speaker who is replete with original and titillating brain children derived from his many years of teaching and practical application of the techniques of personnel selection: Professor William A. Hunt, Chairman of the Department of Psychology, Northwestern University.

Professor Hunt received his AB from Dartmouth and his Masters and Doctor of Philosophy from Harvard. He has served on the faculty of Dartmouth, Connecticut College, and Wheaton College.

Dr. Hunt is a retired Naval Captain, MSC, USNR, having served actively from 1941 to 1946. He is presently a Consultant to the Surgeons General of the Army and the Navy and was on the Army Science Advisory Panel in 1956. It gives me great pleasure to present your distinguished colleague, Professor Hunt.

KEYNOTE ADDRESS: SOME COMMENTS ON MILITARY SELECTION

**William A. Hunt, Ph. D.
Northwestern University**

It was twenty-six years ago today that I reported on duty at the Naval Training Station at Newport, Rhode Island. The pre-war Navy was a formal Navy. I had on my gloves and I wore my dress sword. Somebody forgot to tell me that you didn't have to have it on when you drove your car. As a result, I am one of the few Naval officers who can drive a car wearing a dress sword. As you know, urinary frequency is a symptom of anxiety and when I stopped for gasoline just outside the town of Newport it was the first day that this particular gasoline station had been opened for the season. I went to the men's room and when I attempted to leave I found that the lock was stuck. The window was still barred for winter with very firm wooden bars which were nailed in. There I was, due to report in about ten minutes and locked in

the men's room of a gas station. I finally got out and got to report.

I had a delightful time. Since the Bureau had not as yet decided what insignia an HVS officer should wear, my sleeve was bare except for two stripes. I sat in the Commanding Officer's anteroom for about forty-five minutes while he got a number of the senior officers in the area to come into his office and look through the peephole to see what strange kind of bird had reported without insignia.

They were very kind to a little college professor and, in the middle of the afternoon, I repaired to the Viking Hotel, took a little nap, and then greeted Dr. Cecil Wittson, the psychiatrist, who had preceded me on duty. We had dinner together, went up to the room and, over a bottle of scotch, a bright dream was born. It included the ideal of a computer program into which personality data, job description data, and an overall limiting system of manpower requirements would be fed. Out of this, Gentlemen, would come selection for the Navy.

I am here this morning to ask what happened to that bright dream of twenty-six years ago? Where did it go and why? Looking back now I am horrified by our naivete. Our concept of personality was too simple, our faith in our measuring instruments too great; and our knowledge of war time service criteria was nil. Nevertheless, military selection worked. In the light of today's knowledge we can well be dissatisfied with its efficiency, but considering the state of the art twenty-six years ago, I feel that none of us, in any area of selection, has need to feel ashamed of our emergency contribution at that time.

These were matters of professional naivete. We have made great progress in overcoming it. We were also guilty of social naivete, however, and I think there has been less subsequent progress in overcoming this. Most of us were highly trained professional people with a research background. We had, and still have, three great blind spots concerning our application of selected techniques within the military establishment. We forgot that we had to make ourselves under-

stood in an alien culture and that we had to communicate with our non-professional military colleagues in terms that they could understand and accept. Secondly, it never occurred to us that selection might be applied to ourselves, and that military selection itself was a special task with job requirements that demanded personality characteristics and adjustment quite different from those that had contributed to our civilian success. Finally, we overlooked the fact that the acceptance of any selection procedure applied to society must depend upon its relationship to an exceedingly complex set of social and political pressures, many of which lie outside the scientific goals and techniques of selection per se.

The problem of professional communication with non-professional colleagues is a universal one for the scientist. It demands that the scientist assume the responsibility for his communications and that he make his findings relevant to the value systems of his audience. This communication function demands, often, a fair amount of common sense. The selection literature is replete with examples of the poor judgment I'm talking about. Frequent instances occur in screening in connection with security checks, for example, when qualified personnel are rejected because of the presence of traits or habits totally unrelated to the security problem. I recall one specific case taken from "real life," in which a well-meaning agency investigator was seriously considering refusing clearance to an otherwise qualified twenty-five year old, single man because he was known to engage in a quite common autoerotic practice. The crucial question is, what did this behavior have to do with his being a security risk? As a rule, what a person does by himself rarely ever comes out in a court of law, so in this case cited here as an example, one screening criterion at least had little, if anything, to do with the purpose of the screening procedure. In short, we need a close marriage between the specific screening criteria in force and the mission or purpose of the total selection program.

The people who research, engineer and execute the selection procedure are human

beings as well as cogs in an impersonal system. If you look at society as it has grown and expanded in the last fifty years, I don't think you can come up with any other conclusion than that more social control is necessary. As control becomes more and more necessary, selection is going to become one of the really vital areas of human existence. I like to think of us selectors as the "sorters," both "sorter-inners" and "sorter-outers." It sobered me to realize that we will serve both as judges and court of appeals. We'll make the selection. We'll make the quality control decisions as well as the subsequent evaluation. It is an important fact that we have to face. I listened to you people yesterday afternoon in the room, again in the lounge, and again at dinner. Over and over one of the problems that came up was the personal turmoil involved and the occasional inability to face the fact that in the military you have to make decisions and do things which involve human life. If you get into selection in industry, efficiency is certainly one criterion, but the ultimate criterion in non-military selection seems to me to be social approval. This is not true in a military organization. National survival is the ultimate criterion, and national survival involves a much more severe penalty for some people than it does for others; therefore, when the veneer is taken off selection, the real guts of the problem is left in the hands of you people who have to make some pretty important decisions. I think you all know what I'm talking about. Decisions concerning life and death are difficult for those with an essentially academic or industrial background. It's been amazing to me in the last year or two to see, in a VA hospital which I'm very close to, the emotional problems which the chief psychologist had to go through when he was put on a board to select those people who shall have the advantage and the use of the hospital's artificial kidney. Picture if you will a situation where you are in a position to say what man will live and what man will die. This man was facing this for the first time, and facing it with some difficulty. For many of you it is not a new problem.

Selection is, in my opinion, the most unpopular damned business in the world. This

is in part so because it is an abrogation of the democratic dream. It denies equality of opportunity. We rationalize this by saying while the process denies equality of opportunity, it does open up the possibility of realizing one's full, personal potentiality, or something of this sort. Look at it any way you want, however, some people fall on the wrong side of the fence and some people fall on the right side. For example, we want the best schools for the best people. As a college professor who has served some time on admissions boards, one of the rougher experiences I've had is when one of my colleagues comes up with the problem that his son has been denied a scholarship, or fellowship, or admission. Suddenly I have to bear the brunt of a whole vicious attack on the selection system because it hits him right where it hurts, with his own child. Each of you has your own way of resisting these pressures. I think we make the mistake sometimes of lying down and letting this criticism walk over us instead of fighting back. I see the invasion of privacy issue as the kind of difficulty we've got to expect. As social organizations become tighter and tighter, and selection becomes more and more prominent in the total process, more and more outbursts of resistance to the selection movement will occur. What we've done perhaps is to lie down before the onslaught and not fight back.

A related problem has to do with the restrictions placed upon the use of experimental animals in scientific laboratories. I don't know whether you realize the turmoil this has produced in academic circles. I think we made a mistake in the way we fought this. I think what we should have done is to have gone into the Congress and supported this bill suggesting that we really needed to recognize the value of these animals' lives, moreover suggesting that a rider be attached to the bill extending its provisions to private pets. Some of these provisions were, by the way, that a veterinarian be in attendance once or twice a week, and regular inspection of feeding and housing facilities.

To return to the problem of the invasion of privacy, we might meet this with an ag-

gressive counter attack. We do invade privacy; this is a basic principle of social control. However, to imply that testing is going to invade privacy that has never been invaded before is absurd. Let me give some examples. There is no more intimate aspect of a person's life than his sex life. Sex life unfortunately is connected with venereal disease. If you get venereal disease, this is supposed to be reported and, as a result, the diseased person is prohibited from food handling. The same is true of tuberculosis. Likewise, nobody wants to let the blind drive automobiles, or, though few people ride trains anymore, you certainly wouldn't want to ride with a color-blind engineer. What I think we should have done when the invasion of privacy matter began to come to the forefront was to point out the importance of controls all through our society rather than saying, "Good heavens, there's no invasion."

This gets us into the social and political context within which selection is practiced. In this context, I am reminded of one of our problems toward the end of World War II. We estimated that about five per cent of the Navy was inefficient in performance. These were not psychiatric cases, but ineffective performers in the line of duty. If we could get rid of that five per cent, the performance of the remaining Navy might go up ten per cent; so we would have a ten per cent better Navy with five per cent fewer men. We pointed out over and over again the possibility of inaptitude discharges to handle the situation. A commanding officer has the right to get rid of somebody who is not doing his job, etc. Much to our surprise we found out we couldn't do this. Why? Because if you send Johnny home, and he says, "They threw me out of the Navy because I wasn't doing my work right," and there's a war on, and Mrs. Jones' little boy Jimmy next door is still in there fighting, you produce a tremendous amount of civilian unrest. You can't do this. Yet you have to handle these people somehow. Our problem had an unsuspected public relations aspect to it, but we found a way to handle it. All the inefficient people who really belonged to the Bureau of Personnel were now turned over

to the Medical Department, to Psychiatry, for a new diagnosis, a diagnosis that is an interesting one for the medical profession, namely, "No disease, temperamental unsuitability." This simply meant that the man couldn't get along. The only way we could get rid of these inefficient people efficiently was to declare them a medical problem. So people went home with a medical discharge simply because the public would accept this. It is unacceptable in our culture to make a sick man fight. There are many ways to solve a problem.

In saying that selection always operates in a political context, I am using political in a broad, social sense. Thus the induction of the so-called "Group IVs," or marginal recruits is not a matter of pure necessity or of manpower demands, but largely motivated by a desire to utilize military service as a means of rehabilitation. This is the old argument as to whether the primary function of the armed services is the military defense of the country or rather to be a broad instrument of social policy, available as resource for the medical, educational, emotional, and economic rehabilitation of the marginally equipped citizen.

I am reminded of an argument Dr. Eli Ginzberg and I had at the time of the Korean difficulties. Ginzberg and Hunt were miles apart on the question of who should get into the Navy. The discourse went something like this: Ginzberg looked at me and asked, "Bill, how many people do you think are illiterate in the United States?" I said, "On the basis of our Training Station figures, I'd say about fifteen per cent couldn't read adequately, say at the third grade level." He replied, "That is a little high; I'd say twelve per cent." Well, I said, "I think I'm a little closer to reality than you are, Eli. I'm not relying on census figures. Our people are actually examining these people as they come in from the induction centers. In any case, it is a large number." Ginzberg answered, "Don't you want to do something about it?" I said, "I certainly do, I'd like to put more money into education." "You've got the wrong idea," he said. "The reason these people can't read and write in this land of

opportunity is that they don't want to read and write, and the only way you can ever teach them to read and write is to get them into the military service and make them learn to read and write."

This approach frightened me and gave rise to a recurrent fantasy. I was back in uniform. Russian planes were coming over and I was out in a schoolyard holding my hand up with a whole school building full of Navy recruits back of me, and I was shouting, "You can't do this, you can't do this. I'm teaching our Navy to read and write." Fortunately, the fantasy never came true.

Those of you that are sophisticated about the military know full well that this is one of the few large organizations where you get a more or less random sample of the population over which rigid controls are imposed. Moreover you have the ability, with a little rationalization before the Congress and within certain limits, to do what you want with them. You've got to realize that any proponent of a social program such as I've just given as an example is going to look to the military as a possible means of implementation of the program, and this will produce certain pressures on any selection procedure.

Let me shift from discussing the cultural context of selection, and now comment on some technical matters. Our job is essentially a utilitarian one, to devise a selection system whose benefits adequately compensate for its costs. Operational profit and not theoretical perfection is our criterion for success. You remember the song from "Oklahoma," "Everything's up to date in Kansas City—they've gone about as far as they can go." This is the way that I feel about a good many of the refined techniques of item analysis, statistical manipulations, factor analytic approaches and the like, techniques that we rely on today in attempting to improve our selection procedures. I think we have to ask ourselves whether we're going to spend the rest of our lives hacking away at four per cent of the variance, or two per cent of the variance, or one per cent of the variance; or whether we ought not to give up this kind of "polish" and realize that it is not a reasonable goal to attempt to re-

construct the total personality, that we must give up the dream of the perfect personality theory which predicts everything about every individual. I suspect we have been too fond of fixating on the two to four per cent of the variance and overlooking the fifty and sixty per cent of the variance which we tend to take for granted.

One of the things we might do with profit is to turn from the minor intricacies of personality structure and give our attention to some analysis of the various situations to which people are subjected, realizing that behavior at any moment is a function not only of an individual's personality structure but of the situation in which he finds himself. We have done little of this situational analysis (possibly some under the rubric of motivation) but the addition of some classification of situationally determined stimulus elements promises much for improving our prediction of behavior.

Let me take an example from the field of emotion and the threat which produces it and refer you to William James' old example where a bear is chasing a man and the man feels afraid because he is running away. The question to be answered is this: granting the bear sufficient motivation, and greater physique and speed, why has he not caught the man in all these years that he has been chasing him? The answer, as Carroll Pratt revealed it years ago in a class at Harvard, is that the man has dry ground to run on. The bear's failure to catch up depends then not on the bear's physique but on an environmental circumstance, the absence of a dry track, and final performance is an interaction between physique and environment.

If a bear came into this room today, my reaction would be a lot different from what it would be in the classroom. I don't know quite what would happen. In the classroom I would feel responsible for a group of adolescents. Here I think I would be more worried about how I conducted myself as a Naval officer in the face of my colleagues. And then there is the interesting problem of where does military duty end and where do you begin to worry about protecting any females in the room. In other words, we're

men and we have a traditional attitude toward ladies. We're also Naval officers; we're this; we're that. If you go back far enough, we are all our mother's sons and are acting in part as mother would want us to, and the roles differ with the situation. These are the kinds of things that we've got to get into. We need to throw our analysis of different situations into a mix with personality structure per se to get behavioral prediction.

For some years we have been interested in the behavior of the psychopath, the asocial personality. Among other characteristics, there is a great deal of current interest in the fact that he does not seem to respond well to incidental reinforcement in a verbal conditioning situation (i.e., smiling every time he uses a pronoun or verb, etc.). Eysenck has attributed this to some peculiarity of the nervous system, and related it to extroversion. On the other hand, the psychopath seems to perform better in some learning tasks, if you define learning as improving in your ability to draw three-inch lines with reinforcement provided by knowledge of results.

These two sets of findings appear paradoxical, but I think the paradox disappears if you consider that you are measuring learning in two different social situations. In the first, the task is vague and depends on some attention to and cooperation with the experimenter. The reinforcement is unclear and comes from an assumedly hostile source, the experimenter. In the second setting, the task is clear, the reinforcement comes from one's own performance, and improvement means a chance to excel and show off, which the psychopath is not loath to seize. Any assumption that learning depends on some fixed potential that is not sensitive to differing situational pressures leads you astray.

Let me illustrate from my current interest in the field of alcohol and tobacco habituation. If you look at the history of the study of alcoholism, you find it in a sense, rather like the history of personality. It started out by studying metabolic processes. We looked for a tissue need. I suspect some alcoholics are more sensitive to alcohol and do build up a tissue need, but this does not answer all our

problems. From there we went into the psychiatric, the dynamics aspect in which the answer to alcoholism is understanding the man's problems. If you got rid of his tensions, he no longer would be an alcoholic. This accounted for a certain number of alcoholics. From there we went to a cultural approach, assuming it to be a cultural problem. We studied family and nationality patterns of drinking and this gave us further knowledge of the cultural elements that encourage or discourage the use of alcohol, but our answers are still incomplete.

I would suggest what seems to me to be an extension of the situational approach, namely that we forget our concentration on the personality structure of the alcoholic and concentrate on the characteristics of alcohol itself. One might almost say study the ecology of alcohol. For instance, in some form or other it is easily available, relatively inexpensive, the dosage can be controlled, the effects are transitory, a lethal dose is difficult to imbibe, it has social symbolism across a wide range of socio-economic strata, etc. These characteristics suggest its potential for addiction, and such a description of the ecology of the drug might in turn give us new insights into the personality of the user. At the very least it gives us further material for understanding the interaction between drug and user, and should give us a better base of understanding for improving our prediction of behavior.

Along similar lines I've had some interest in the use of barbiturates in suicide. What interests me again is what is a barbiturate? As opposed to alcohol, barbiturates are difficult to obtain. I wonder if this may not have appeal to some people. Once you expend the effort to get them, you have made a commitment. It's like resigning your job after twenty-one years, or joining the church, or suddenly running for office and being elected. These are gestures that commit you to a course of action. I would suggest that if you study the characteristics of barbiturates you then will be led back to the personality of the barbiturate user. I think you will get many answers by studying the characteristics of the drug, the ecology, if you want, of the

situation, and you can begin to see what angles to attack the problem with in your educational programs.

Another profitable approach to improving our selection procedures would seem to me to come from a better understanding of what level of prediction is pertinent to our particular selection task. We are all acquainted with the use of progressive screening procedures going from a first rough screen to later, finer selection measures. Thus during World War II the Navy's psychiatric selection program often used a paper-and-pencil test as a first sorting device for narrowing the original number of recruits to a smaller "pay dirt" population who then received a finer screening through a personal psychiatric interview. The point I wish to make is that we tend to look down our noses at some of these rougher techniques, to overlook their efficiency for certain situations, and to evaluate their validity not in a functional sense in terms of their performance in a specific setting for a certain purpose, but rather against some finer criterion, often a theoretical one, which may not be appropriate for the particular situation in which they are used.

Let me give you a specific example from a study that Cecil Wittson and I did in World War II.* We were in a Training Station and there was a Naval Hospital next door. In those days no Training Station Medical Department could make a diagnosis. This was the function of the Naval Hospital. When we found a psychotic or neurotic, the man was sent over to the Naval Hospital—not with a diagnosis accompanying him, but with the notation, "DU-Observation," which translated, "Diagnosis Unknown," and left the diagnosis up to the hospital staff. Fortunately for research purposes, we had very carefully diagnosed everybody we sent over there. This gave us a chance to measure the reliability of the diagnostic process by comparing their diagnosis with the original one we had made. Now, if you get down to a specific diagnosis, such as psychoneurosis hysteria,

psychoneurosis anxiety, manic psychosis, affective psychosis versus schizophrenia, for example, there is a great deal of disagreement. There was agreement in only thirty per cent of the cases. This is pretty low, and agrees with other studies of the unreliability of psychiatric diagnosis. But Cecil and I thought that there was some value in saying a man is psychotic as opposed to being neurotic or suffering from a character disorder. So we grouped the diagnoses into three categories. It turns out that the agreement rose at most to sixty per cent. Now we are approaching useful reliability.

It is important to realize that this reliability attained by grouping specific categories into one general class, i.e., putting all varieties of psychoses under one general heading, "psychosis," has functional utility. Irrespective of specific differences in behavior, the general characteristic of being psychotic has important implications for patient management, i.e., the need for custodial supervision, the possibility of the patient being a danger to himself or others, difficulties in communication, etc. The same thing is true of neurosis and character disorder. Thus, a disagreement in specific diagnostic category may still imply agreement on a broader level involving behavioral predictions having a genuine social utility.

We then decided that we really didn't give a damn about these scholarly niceties. The thing that we were supposed to do was to pick out people who were not suitable for naval service. If you use this general category, unsuitable for naval service, the agreement between the two groups of psychiatrists was ninety-three per cent, indicating extremely good reliability for the basic purpose of screening. We might differ as to specific etiology, but on the basic judgment of suitability we were in agreement, and this determination of serviceability was the ultimate goal of psychiatric selection.

In the light of this, it has always seemed to me unfair that so many attacks against psychiatric selection are based on the unreliability of the psychiatric diagnosis using the criterion of agreement between specific diagnostic terms, rather than the more sophisti-

*Hunt, W. A., Wittson, C. L. & Hunt, E. B. Theoretical and practical analysis of diagnostic process. In P.H. Hoch & J. Zubin (Eds.) *Current Problems in Psychiatric Diagnosis*. New York: Grune & Stratton, 1953. pp 53-65.

cated (and admittedly more difficult) analysis of the broader implications of the diagnostic procedure as illustrated in our study. It is obvious that while psychiatrists may not be able to agree on the finer behavioral predictions implied in the use of specific diagnostic categories, they could agree on the prediction of suitability for service, which, I submit, was their most important function in the screening program. A diagnosable personality disorder may not receive a psychiatric discharge during subsequent service, but it is significant if he is separated for bad conduct, a psychosomatic complaint, general inefficiency, or in many cases even a dependency discharge, for all of these may have implications for the original diagnosis.

I should like to conclude my remarks this morning on a note of warning. The type of recruit we are getting in the military today is quite different from the type of recruit we got in World War II. The warning is that the present recruit is also different from the recruit you can expect if we shift from peace-

time conditions or those of limited war to an all-out conflict. We are doing a creditable selection job on our present military population, but a total war will present us with a different set of problems and an unselected, or, at the very least, a much less selected population than we are handling at present, and you may be professionally as unprepared for it as we were in World War I and World War II.

The big paradox of research on selection in the military or even on training in the military is that you have to prepare for war in time of peace, and almost every peacetime program you have gets shot down once total war breaks out and you have universal conscription.

You have a royally selected bunch in the Navy now, but God help you when the dregs of the population fall upon you, as they're going to if we ever get into an all-out war. Keep that in the back of your mind and remember that your selection measures have got to fit this kind of situation too.

SESSION I

BRIEF UPDATING OF SELECTION PROGRAMS BY REPRESENTATIVES OF PARTICIPATING ACTIVITIES

SOME BRIEF COMMENTS ON THE SELECTION PROGRAM AT THE NAVAL AEROSPACE MEDICAL INSTITUTE PENSACOLA

Rosalie K. Ambler

USN Aerospace Medical Institute, Pensacola

In Pensacola our attitude toward the selection process and selection research has evolved from our experience with it. Early in this experience we learned that we could improve on primary selection if we considered selection as a continuous process. That is, we added training data to the residual validity from the primary selection variables and profited from a secondary selection system. Acquisition of a computer a few years ago gave considerable impetus to this effort, and we now have a quality control system within aviation training that I assume is familiar to most of you.

It has also become obvious to us from discussions within the training Command and the Combat Replacement Air Wings (CRAWs) that performance in the CRAWs depends on optimal standards in training. And we also expect to find a relationship between CRAW performance and fleet performance. In effect, prediction of performance in each succeeding phase in a training and/or operational career should rely on prediction variables from all preceding phases. Fleet performance, therefore, depends on how good the selection process works at the procurement level, at the primary selection level, at the classification level, at the secondary selection, and at all training levels.

The following figure presents a model of our approach. From an examination of this figure it is seen that selection methodology involves a data bank that must be supplied from variables taken from across the entire career spectrum. There are at least two research foci leading from this model, both of which are receiving considerable emphasis in the present program at Pensacola.

1. We are developing a method of computerized construction of paper and pencil tests that will utilize an "item bank." The idea of an "item bank" is not new of course, but the problem in the past has been in keeping item banks updated. We think that computer technology will enable us to overcome this problem and in so doing maximize validity for gross screening devices at the primary level.

2. A major effort at Pensacola is centered around the development of a dynamic testing environment where the subject is placed in a test situation on-line with the computer. Material can be programmed to the subject in a manner that will permit a given test stimulus to be influenced by the previous response of the subject. A broad range of test content will be possible in the test booth as it is now conceived. At this time it is envisioned for use at the secondary selection level.

Very briefly, these statements, together with the Selection Model presented in the accompanying Figure provide some idea of the nature of the Selection program at Pensacola at the present time.

Discussion Following Miss Ambler's Presentation

CAPT CHRISTY: It occurs to me that some of the work that Dr. Rahe is doing at San Diego on life incidents might have a lot to do with aircraft accidents, which might be one criterion to aim for. In evaluating a pilot's performance, one needs to take into consideration the total situation, for example, he flies differently if he has a wife and kids. One thing that has troubled up about our selection procedures is that they emphasize measurement of ability to get through flight training, but at the same time, the same procedures are not ordinarily highly predictive of ability to be a good combat aviator. In other words, our selection procedures cannot, at the training level at least, discriminate between a Pappy

PRE-TRAINING VARIABLES



TRAINING VARIABLES

PRIMARY SELECTION

Intelligence Interests
Aptitudes Motivations
Abilities Anthropometrics
Personality Medical History
Personal History Physiological

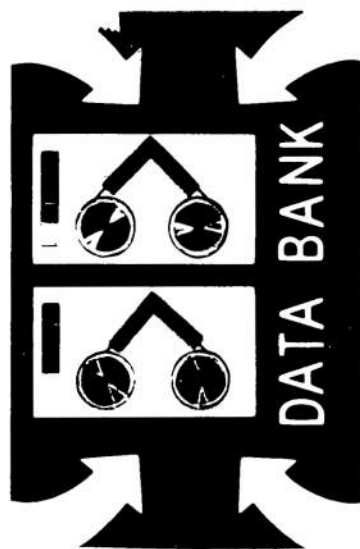
SECONDARY SELECTION

Special Abilities Health Record
Performance Peer Ratings
 academic military
 flight disciplinary

Training Pipeline
↓

OPERATIONAL (CRAW)

— Performance



CAREER VARIABLES

PROCUREMENT

Age
Education — level
 kind
Draft
Marital
Socio — Economic
Ecologic

Accidents
Health Record
Fitness Reports
Changes of Designator
Retention

Boyington and a Joe Foss on one hand, and the run-of-the-mill pilot on the other. Now I wonder if you have been able to get some feedback from carrier groups as to who does well in combat, possibly from peer rating of fellow pilots? These data, I should think, could be fed into your computer model quite well.

MISS AMBLER: We want to do this. Every time anyone talks about combat criteria, peer ratings emerge as the most desirable kind of data to get. There have been some preliminary efforts made to get data of this kind in Vietnam. However, we have not yet gotten far. Work is in progress.

CAPT CHRISTY: These kinds of data also become complicated by morale, leadership, and other factors which are involved in the individual's ability to excel.

MISS AMBLER: There are other complications. You may remember from World War II that the pilots were often reluctant to provide us with peer data. Too much time was involved. Also, concern about the disclosure of peer interrelationships is probably involved here. We presently have a project underway to provide combat performance criteria to be fed into the system. We quite frankly, do not know whether the flight aptitude ratings are correlated with combat success or not.

DR. GELL: In 1939, I was a shipmate in San Diego with Boyington, Joe Foss, and another incorrigible named Miller. It seemed that these men did not adjust too well to the peacetime flying Navy, as I remember their pre-war service reputations. The subsequent wartime record of men like Foss, Boyington and Miller are matters of record. They became aces several times. Miller got to be known as a one-man task force for such things as bombing the island of Truk alone in his Liberator with no fighter protection. These men were exceptions to the rule, but they certainly exist. It seems that there are certain personality traits that don't emerge in a nice quiet atmosphere. I'm amused by Captain Christy's remark about pilots with a family. When I went through flight training I was constantly being bedeviled by one

of my instructors because I wouldn't shoot for a power landing in a circle at 55 knots. I always came in at about 60. Of course, the stalling speed of the airplane was 47 knots. He asked me one day point blank what the hell was the matter. I said, "Well, I'm simply adding 5 more knots for my wife and two kids."

CAPT CHRISTY: The comment made a moment ago about the difficulty of getting peer data in a combat zone is true. However, if you collect these data when the carrier starts back, the pilots have time on their hands before their return to the states. We've done this with some success.

DR. HUNT: To look into the future, have you given any thought to some kind of control on what the use of the data bank is going to mean to the man's performance? The system is going to be known; he's going to be aware of it.

MISS AMBLER: Yes, they are aware of our quality control system now. I do not know of any effect on their performance in training resulting from the system.

CAPT CHRISTY: You mean the quoting of the information — matters of its confidentiality?

DR. HUNT: No, the use of it. I think the Internal Revenue Service should have started this years ago. I'm serious about this. What are effects of an income bank which they're developing? These data banks are developing in all fields.

CAPT CHRISTY: This point came up in the National Mental Health Council several weeks ago when the possible effects of a nation-wide bank of psychiatric histories were questioned. For example, the bank might contain a history of some sort of a breakdown a person had at the age of 25 who, now at age 40, is running for a political office.

How to maintain the confidentiality and yet use the information constructively is a most crucial question.

DR. WHERRY: One of the things that we had planned doing with the data bank at

Pensacola was to use the information in it to determine which of the various pipelines the men would go into, for example into jets, multi-engines or helicopters. We thought that we possibly could discover certain of the skills predictive of which of the pipelines is most appropriate for each man. This information could be put into the system for each man prior to his stating a preference. At present a lot of them think they've got to say that they want jets because that's the thing to say. However, it is possible to sway his preference to small planes, and possibly to multi-engined planes if you have the data to show he would make out better in one kind as compared to another.

DR. HUNT: At the risk of being risqué, but to emphasize the point strongly, let's face the fact that we're moving into the kind of world in which data banks of all sorts will be commonplace. Soon, before going to bed with your wife you'll pull the shades up instead of down. Privacy is gone.

CAPT CHRISTY: During the war, for example, we had this problem. Who would go multi-engines and who fighters and dive bombers. Not having good criteria, we used such information as "did he or didn't he get the bends during four altitude runs." If he did, he was out. But this criterion broke down completely. Hopefully, better criteria for such selection decisions are now available.

DR. WEYBREW: Do your instructors have access to the information as to what we in submarines call "marginals"—those people on the borderline? If the instructors know who is marginal, I'd think you would bias the whole "pipeline." If you put a tag or a label on a person, your instructors are going to watch for inappropriate or ineffective behavior, thus increasing the odds that the "marked" men will indeed fail the program.

MISS AMBLER: The way the present system operates, the instructor is really the one who initiates the marginal label, presumably from his observations during training. Then

they come to us through the pilot disposition board for a prediction as to the probable outcome of the student in question. At a given point in training, the data that go into this prediction are everything that we have in our data bank. In short, the instructor initiates this tag rather than our doing it.

DR. HUNT: Isn't it the whole basis for mental health in the schools these days, that the teacher knows the peculiarities of the pupils?

DR. WEYBREW: But in Pensacola you have a college-caliber man who wants to get a commission. Once he is labeled as marginal, regardless of who does the labeling, I would think your instructors are going to look for these faults and probably find them.

MISS AMBLER: Yes, we have some of these in the NFO School, the non-pilot school. We have had requests from the staff to flag their students. We have not yet acquiesced to this.

DR. WHERRY: At the command, the man's jacket follows him around all through training, so it isn't much of a point since he has access to all this information anyway. We have a feeling that the instructors are putting the information together prior to letting the computer integrate the various scores.

MISS AMBLER: I'm not sure that what we're doing here with this model is conceptually a whole lot different from what we've been doing all along. It's just that we are able to do it in a far more systematic way and on a much larger scale with the computer.

DR. HUNT: A lot of success that astounded our academic colleagues in selection during World War II was due to the fact that most of the kids thought the FBI had all the data on them anyway, so they answered every personality test item more or less honestly. We handled the interviews this way too. This results in a completely different climate in your selection program.

**NAVAL RECRUIT SELECTION
PROGRAM AT THE
NEUROPSYCHIATRIC RESEARCH UNIT,
SAN DIEGO**

**John A. Plag, PhD
Navy Medical Neuropsychiatric
Research Unit
San Diego, California**

Since the beginning of World War II, the Navy's Bureau of Medicine and Surgery has operated psychiatric evaluation units at the various Navy and Marine Corps recruit training installations. The major mission of these screening facilities has been to identify new enlistees who are neuropsychiatrically unfit for service. Ostensibly, the purpose of early identification and separation of such enlistees from recruit training has been to reduce subsequent service attrition and improve the military effectiveness of personnel assigned to fleet operating units.

For a variety of reasons, such as continued high rates of attrition among personnel in the fleet, evidence from studies contrasting the validity of clinical and actuarial prediction, etc., there was cause for questioning the value of psychiatric screening procedures at recruit training commands. As a result, in 1960, the Bureau of Medicine and Surgery directed the Neuropsychiatric Research Unit in San Diego to conduct a prospective study of the preventive psychiatry program at recruit training commands. The following report is a brief summary of a series of separate studies which have been conducted in the past or which are presently being conducted in conjunction with this research program.

The Research Data

Eleven-thousand enlistees who entered the Navy at the two recruit training centers at Great Lakes and San Diego during four seasons of the year in 1960 and early 1961 comprised the research sample.

Predictor and criterion data were collected for the sample subjects throughout their first enlistments, which, for most, terminated in 1964. For the majority of the studies conducted on the basis of these data, the predictors were related to a criterion termed

"naval effectiveness." Effective sailors have been defined as those who complete their tours of duty and are recommended for reenlistment. On the other hand, non-effective enlistees are those who require separation from the Navy prior to the expiration of their enlistments or ones not recommended for reenlistment by their commanding officers. For the sample of 11,000 enlistees who entered service in 1960, approximately 28 per cent were classified as being non-effective.

Other criterion data, such as two-year fleet-performance measures, in the form of division officer ratings of adjustment, record of disciplinary action, semi-annual marks, and pay grade were collected for all subjects. Finally, physical and psychiatric illness data, like number of admissions to the sick list, number of days hospitalized, and diagnosis were obtained for some of the sample subjects throughout their first enlistments.

The Validity of Psychiatric Screening

As practiced in 1960, two major elements comprised the process of psychiatric assessment at the recruit training commands. The first was the initial screening interview — a brief, one to two-minute clinical examination of all recruits entering training. On the basis of this examination, subjects who were suspected of possessing emotional and/or intellectual abnormalities severe enough to preclude their military effectiveness were designated for a trial of duty and given a more thorough follow-up evaluation after several weeks of training. Recruits who, on the basis of these follow-up evaluations, were considered to represent psychiatric liabilities were subsequently brought before aptitude boards and usually discharged from service. During the 1950's recruits discharged by aptitude boards ranged between three and six per cent of enlistee input. The decision to separate enlistees from service constituted the second major element in the screening process.

Evidence from the research data has suggested that neither of these two clinical decisions is as valid as had once been assumed. When related to four-year effectiveness, the

correlation of the initial screening interview was found to be only .19, a value which is of course statistically significant, but of low magnitude. A considerable number of the classification and demographic variables for these subjects were found to have significantly higher correlations with the effectiveness criterion (e.g. education = .32; GCT score = .24).

Experimental procedures utilized in conducting the research made it possible to study the fleet adjustment of a group of 134 enlistees who were purposely retained in the Navy despite decisions of recruit aptitude boards that they be discharged as temperamentally unsuitable. While some significant differences on criteria of fleet adjustment and performance were found between these subjects and those in a matched control group, it was determined that approximately 50 per cent of the subjects presented to aptitude boards not only completed their tours of active, obligated duty but were recommended for reenlistment by their commanding officers. These results warrant the conclusion that the psychiatric screening of recruits can be practiced too enthusiastically.

Actuarial Prediction

Because the research findings suggested that the clinical prediction of four-year military effectiveness for recruits was less than satisfactory, attention was directed towards the identification of demographic, personal history, and initial performance characteristics which might be predictively useful. Actuarial prediction for the purpose of per-

sonnel selection was conceptualized as an ongoing, multi-stage process rather than a static, single-prediction phenomenon. In other words, predictions of eventual effectiveness can be made at any point during enlistees' tours of active duty, with various background and early performance characteristics having different unique predictability depending upon the time that such predictions are made.

In the research investigation, the validity of actuarial predictions was evaluated at five stages preceding and during enlistees' first tours of duty. These stages were:

(1) Recruiting office, (2) Armed Forces Examining and Entrance Station, (3) Second week of recruit training, (4) Final (ninth) week of recruit training, and (5) after two years of active duty. Table 1 shows the variables which were added to the analysis at each of the five prediction stages.

The derivation of each prediction formula was accomplished through the use of multiple regression technique, preceded by exhaustive studies of the linearity of predictor-criterion relationships and the unique validity of variable interactions. At each stage in the analysis, the derived predictor composite was cross-validated and tables were constructed to show the probability of four-year effectiveness for different combinations of the predictor variables.

Table 2 shows the cross-validity of the derived predictor composite at each stage of the analysis. For the sake of brevity, individual item validities are not reported here.

Table 1
The Prediction of Four-Year Effectiveness for Naval Enlistees:
Variables Added to Regression Analysis at Each Prediction Stage

Stage I Recruiting Office	Stage II AFES	Stage III Recruit Training Second Week	Stage IV Recruit Training Final Week	Stage V Two Years Active Duty
Age	Armed Forces Qualification Test	General Classifi- cation Test	Average Weekly Test Score	Division Officer Rating
Education	Active Duty Obligation	Arithmetic Test	Disciplinary Action	Pay Grade
Birthplace	Recruiting Area	Mechanical Test	No. of Recruit Training Trans- fers	Disciplinary Action
No. of School Grades Failed	Prior Service Rejection	Clerical Test	Duty Assign- ment	Semi-Annual Marks
No. of School Expulsions		Psychiatric Interview	Peer Nomina- tions I	
Family Stability		Peer Nomina- tions I	Peer Nomina- tions II	
No. of arrests				
Race				
Religion				
No. of Siblings				

Table 2

The Cross-Validity of Variable Composites Used for Predicting Four-Year Naval Effectiveness at Different Stages of Enlistment and Training.

Stage	Description	Cross-Validity (r)
1	Recruiting Office	.36
2	Armed Forces Examining and Entrance Station	.38
3	Recruit Training-Second Week	.40
4	Recruit Training-Final Week	.42
5	After Two Yrs. of Active Duty	.52

It is of interest, however, to mention some of the variables which make major contributions to the multiple prediction. Level of schooling prior to enlistment and number of school expulsions are two such variables. The former contributes uniquely to the forecasting of four-year effectiveness at each of the first four prediction stages, while the latter is uniquely related at all five prediction stages. At the two prediction stages in recruit training, peer nominations were found to be highly correlated with the criterion; and in stage five all of the two year measures were uniquely related to four-year effectiveness.

The results of this study indicate that actuarial data are far more valid for predicting military effectiveness than are clinical judgments. Tables of effectiveness probabilities, which have been prepared from the research data, have a wide range of applicability as an aid in the making of decisions regarding the enlistment, retention, and assignment of enlisted personnel. In addition, they can be used as an index of the quality of personnel entering the service or entering specific military assignments. On the other hand, because of significant shifts which occur from time to time in the characteristics of newly enlisted personnel, and because of the existence of considerable error in the derived probabilities, it would be unwise for these findings to be applied indiscriminately and without the exercise of due caution.

Psychiatric Illness

While military effectiveness, as defined in this study, is certainly an appropriate criterion for validating clinical and actuarial

predictions, admission to a hospital because of psychiatric illness constitutes another highly relevant measure of enlistee adaptation. What is the relation between military effectiveness and psychiatric illness? What is the incidence of psychiatric disorders among first-term enlistees? Are the pre-enlistment characteristics of enlistees which are predictive of military effectiveness also predictive of psychiatric disability? Answers to these and other similar questions were obtained through a review of the medical records of all 11,000 sample subjects following termination of their first tours of active duty.

It was found that five and a half per cent of the population at risk was admitted to the sick list for psychiatric conditions at some time during the period of their first enlistment; and another 3.2 per cent was administratively discharged from service because of diagnosed emotional pathology without admission to the sick list.

On the average, psychiatric patients completed 435 days of active-duty service prior to their hospital admissions and remained on the psychiatric sick-list for approximately thirty days. Seventy-five per cent of the psychiatric patients were diagnosed as character and behavior disorders.

The disposition and eventual military effectiveness of the hospitalized subjects is of particular interest. Only twenty-three per cent were returned to duty and judged to be free of any disqualifying defect, while the remainder were discharged from service. Of those returned to duty, approximately 58 per cent were either subsequently discharged from service prior to the completion of their active obligated duty or were not recommended for reenlistment by their commanding officers. On the whole, then, first-term enlistees who are admitted to the psychiatric sick-list have only one chance in ten of rendering effective military service.

When compared to a group of control subjects, psychiatric patients were found to have experienced more physical illness, to have achieved fewer promotions in rate, and to be overrepresented in the hospital corps-

men specialty. A variety of characteristics, including level of schooling, Armed Forces Qualification Test score, family stability, number of expulsions from school, reason for enlistment, etc., were identified as being significantly predictive of psychiatric hospitalization. Most of these characteristics are identical to those which were found to be predictive of general non-effectiveness.

Other Studies

During the past year, at the request of the Department of Defense, an evaluation has been made of the adjustment and performance of enlistees scoring in Category IV on the Armed Forces Qualification Test. Twelve hundred and sixty such subjects, who were part of the 11,000 sample enlistees entering service in 1960, were compared with 500 average (AFQT score of 50) sailors on a variety of criterion measures. Category IV enlistees were found to score lower on measures of military performance and to have appreciably lower rates of overall naval effectiveness, but to be similar to average enlistees with respect to discipline and illness rates. Four pre-enlistment characteristics were found to be valid for predicting four-year naval effectiveness among Category IV personnel. These variables were years of schooling completed, number of school expulsions, number of arrests, and AFQT score itself.

In another study, an examination of rates of effectiveness for different occupational specialties has revealed that hospital corpsmen have significantly more attrition than other ratings. For this specialty, however, attrition appears to be highly variable depending upon the type of duty to which corpsmen are assigned following graduation from Corps School. For example, those assigned for duty with the Fleet Marine Force are more effective than are their counterparts who are assigned to shore-based installations. These highly interesting findings have demonstrated the need for a longitudinal investigation of personal history and environmental factors which may be related to corpsman adaptation. Such a study is presently being designed.

Additional ongoing investigations include an epidemiological study of physical illness among first-term enlistees, the differential prediction of various types of naval non-effectiveness, factors associated with the fleet adaptation of enlistees returned to duty from the psychiatric sick list, and the eventual adjustment of enlistees who require assignment to special training companies in recruit training.

A study of the process involved in the initial modification of behavior among newly recruited enlistees is another major research undertaking which has direct relevance for the Navy's preventive psychiatry program at recruit training commands. The identification of changes in affects and attitudes as a function of the training process and an evaluation of specific techniques for enhancing enlistee adjustment are substudies planned within the framework of this project.

Discussion Following Dr. Plag's Presentation

DR. RIMLAND: Do you have any data on whether recruiting personnel are using the actuarial tables that you have developed, or do they just ignore them and fill their quotas as best they can?

DR. PLAG: No, I don't know whether they are using them or not.

CAPT CHRISTY: As far as I can tell, they are not. They have their quotas and it doesn't make any difference whether the man has zero probability of succeeding—if they are five men short, they will go ahead and recruit him anyway.

We've been trying to figure some way to put a penalty on them for this, arguing that if half the men they recruit "bust out," they will have to recruit 50% more the next month in order to maintain quality. But so far we haven't been able to come to any agreement with the recruiters on these points. One thing that Dr. Plag mentioned is that judgments for some of these very low-caliber people were made by the psychiatrists as to those who would and those who would not "make it." As I remember, 67% of the former group made it and only 48% of the others made it, or something about like that. But, only one person out of the whole group made a rate.

Great Lakes has some relevant data on population samples with GCT's of 40 and below. Their findings show that it takes 10 to 15 times as many men to get a rated man out of this group in two years as compared to the normal enlisted population. Facts like these must impress the recruiters, but so far we're not too hopeful.

DR. RIMLAND: Do you have any data on those who might be classified as "effective" and who are requested to remain in the Navy but choose not to?

DR. PLAG: We have the information but we haven't analyzed it. It would be very important to determine who reenlists. What are the relationships between some of these characteristics I've mentioned and reenlistment rate? I don't know, but we do have the data on hand.

CAPT CHRISTY: I think one important concept involved here is whether the man has a soft skill or a hard skill. Industry will grab the hard skill in most cases. I wonder if you have any comments now with regard to whether the recruiters are, in fact, scraping the bottom of the barrel or whether they actually can be quite selective in picking and rejecting among the available Group IV's.

DR. PLAG: I talked with the Head of the Recruiting Office in Los Angeles last February. I was apprised of the fact that about 15% of the Navy enlistees must of necessity come from men with AFQT scores between 30 and 10. He told me that despite this fact they are still able to accept most of the men who have finished high school and those with no serious arrest history, and ones who are 18 or older. You will recall the age variable as being one of our best predictors. Right now I've been told that the mean GCT score of men coming into the Recruit Training Centers is around 53; the percentage of them who have finished high school is around 90—it varies between 95 and 98%. This is excellent recruitment, I would think, particularly as compared to our 1960 sample. Around 57% of those finished high school. So they are not scraping the bottom of the barrel by any means.

DR. RIMLAND: You ought to be quite worried about the Group IV data that you have because DOD is really counting on the Navy to take more Group IV's all the time. I would think the recruiters have been quite selective because a lot more Group IV's apply for the Navy than the Navy can accept and they want to avoid loading the jungles of Vietnam with this caliber of man. In the Army the Group IV's represent a much poorer group. It appears to me that someone looking at the Navy's Group IV data without realizing the selectivity that has been exercised in this sampling might come to some inaccurate and rather dangerous conclusions.

CAPT. CHRISTY: There is another DOD control on how many enlisted men you can take in the GCT range of high 50's or 60's. Some of these men with high test scores wait between 4 and 5 months before they can be taken in under the Navy's quota.

DR. WISKOFF: I just want to mention a couple of things since I'm from The Bureau of Personnel. We know a little about what is going on now as far as selecting from Group IV's. In the total distribution Dr. Plag is right in that the mean GCT and ARI of the input into the Navy is fairly high, but we have a very peculiar kind of distribution coming in now. Something like a bi-modal effect results when we are forced to take 15% Group IV's. But by the same token we have pretty good selectivity because of the draft and so we are able also to take Group I's and Group II's. We are getting in IV's, I's, and II's, but peculiarly, few III's. As far as the quality and selectivity of recruiters goes there is something coming up shortly to increase the selectivity. One of the problems the RTC's are experiencing right now has to do with the literacy of the Group IV's. There were plans to re-institute recruit preparatory training as they had it back in the early 1950's. This is a thirteen-week or more course. I think this is going to come about. I believe that very soon a literacy screening test will be in use in the field. This should make quite a difference in the input population which may be coming in very shortly.

SELECTION AT THE NAVAL ACADEMY
CDR Pat O'Connell, MC, USN
U.S. Naval Academy, Annapolis

Let me preface what I have to say with the statement that what we have at the Naval Academy is really a clinical unit, so I'm afraid I am not going to sound much like a researcher. But let me start out by reading the mission of the Naval Academy:

"To develop midshipmen morally, mentally, and physically and to imbue them with the highest ideals of duty, honor and loyalty in order to provide graduates who are dedicated to a career of naval service and who have potential for future development in mind and character to assume the highest responsibilities of command, citizenship, and government."

A statement about the history of the academy and the Mental Hygiene Unit seems in order at this point. The United States Naval Academy was started in 1845 by Secretary of the Navy George Bancroft. There were something like 9 midshipmen students at that time. The history of the Mental Hygiene Unit began in 1950, not with the unit itself but with one psychologist and one psychiatrist. LT John Conger was the clinical psychologist and a LCDR Wright was the psychiatrist. The psychiatrist was sent officially to the hospital; his billet was there. The psychologist was in Bancroft Hall which is the dormitory for all the midshipmen. Actually they were both working in Bancroft Hall but in separate quarters. In 1953 the Mental Hygiene Unit was established and in 1955 the psychiatrist's billet at the Naval Academy was made official. In 1956 a third billet for a clinical psychologist was added. This comprised the facilities and the personnel of the Mental Hygiene Unit, one psychiatrist, two clinical psychologists. About six years ago the inside of Bancroft Hall was rebuilt, resulting in a consolidation of the Medical Department and the facilities of the Mental Hygiene Unit. Prior to that time the psychologists were up in the front part of Bancroft Hall; the psychiatrist was down in the Medical Department.

So at present we are all together in quite modern facilities that are small yet adequate.

We all do the same sorts of things. As I said before, we are a clinical unit. We spend most of our time in clinical work and therefore have little time for research or thinking about research problems.

The size of the classes of midshipmen totals about 4,000. The 850-900 fourth year classmen come from a freshman class of about 1300. On the average 31% will drop out for one reason or another during their four year stay. This average has been fairly steady over the past ten years. One quarter of this 31% will drop out during the plebe summer. By the end of the first year 2/3's will be gone. By the end of the second year which is their "third class year" 90% will be gone. And by the end of the third year 99% of this 31% attrition will have left the academy.

Now here are some figures which, incidentally, are about 10 years old, yet they indicate what this loss might mean financially. I am sorry that I don't have any current figures, but you may be able to extrapolate these. It has been estimated that on the first day one midshipman cost the government \$3,000. His total training estimate was about \$27,000 at that time. With regard to those who remain and successfully complete the academy program and are commissioned, I have heard estimates that from 60-80% pursue a Naval or Marine Corps career of at least 20 years. I would think that this speaks well of the selection for and the program of the academy.

Now, how are these classes chosen? Selection at the Naval Academy has two aspects, the positive and the negative. Selection from a positive point of view is carried out by the Office of the Dean of Admissions. Dr. William S. Shields, Senior Professor and Dean of Admissions, has been associated with the academy for about 25 years. Over the past few years he has been formulating and developing a point system for weighting factors and qualities in prospective midshipmen. This system can be compared with the "risk categories" devised by Newell Berry and John Plag at the Navy Neuropsychiatric Research Center, San Diego, for use in evaluating the potential of Marine recruits. I believe Dr. Shields' System uses currently a total of

80,000 points. Such items as College Board scores, high school average, extracurricular activities, recommendations are all weighted tentatively, and the system is being developed, evaluated, and improved each year via computer as each sample class progresses through the academy. To give you some idea what factors and qualities Dr. Shields considers important, here are some figures from the profile of the Class of 1970, prepared by the Admissions Office. Of 5039 applicants examined, 2180 were found scholastically qualified. Of these, 1339 were appointed and admitted. The mean scores for college boards were: SAT Verbal—586; SAT Math.—655; Eng. Comp.—569; and Math.—642. Seventeen per cent of the class had some college preparation (one-half year or more) prior to admission. Seventy per cent were in the top fifth of their high school class, twenty-one per cent were in the next fifth, six per cent were valedictorians or salutatorians, forty per cent were in the National Honor Society, nineteen per cent in other honor societies, thirty-seven per cent had received academic awards and medals, ten per cent had won Science Fair Awards, sixteen per cent had won National Merit Scholarship recognition, thirty-five per cent had been class or student body officers, sixty-nine per cent had participated in varsity athletics, forty-seven per cent in intramural athletics, forty-seven per cent had been in academic clubs, thirty per cent were in dramatics, thirty-four per cent were in musical groups, and seventy-four per cent were in Youth Service Groups. Now in the system, academic achievements are weighted roughly 75%, and extracurricular achievements 25%. This gives you an idea of the kinds of variables in which Dean Shields has been interested. He feels that the best single predictor is high school rank; and this sounds like what I have heard here this morning. The combined best predictor is high school ranking and college board scores, Math. in particular. Recommendations also appear to be important factors in prediction.

Of course, selections come from many different sources, from Representatives and Senators, Naval Reserve, Marine Reserve,

regular Navy and regular Marine Corps, sons of deceased veterans, and Presidential appointments. Each member of Congress can have, I believe, five men in any one of the service academies at one time. And for each of these positions he has alternates and they may run up to five or six sometimes. So there is quite a bit of choosing to be done.

I mentioned before that the total attrition was 31%. A breakdown of this group shows midshipmen coming from civilian life have an attrition rate of 25-29%. Those coming from the Naval and Marine Corps Reserves have a rate of 37½%. And those coming from the Regular Navy and Marine Corps, on active duty, have the unduly high rate of 46%. These latter are the so-called Napsters, who have first been to the Naval Academy Preparatory School at Bainbridge, Maryland. The academy is quite concerned about their high rate of attrition. One of the reasons for their inability to complete the program seems to be that having been through recruit training, having had active service, occasionally in combat areas, and being older in many cases than their contemporaries and midshipmen superiors, they don't like being told what to do by someone younger than they.

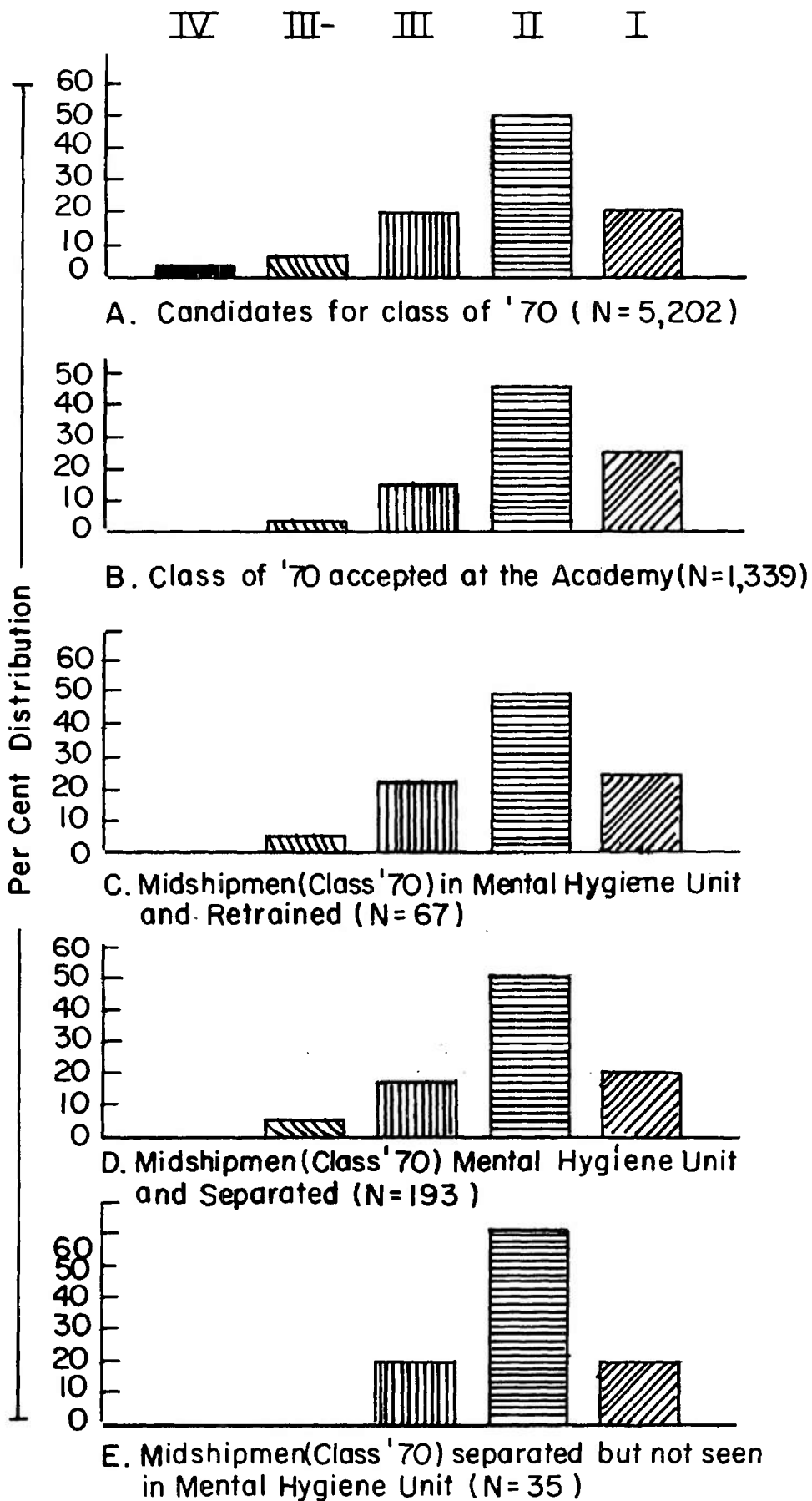
Dean Shields has found that reasons for attrition fall into two broad categories: voluntary resignations (two-thirds), and academic discharges (one-third). Ten years ago these figures were apparently reversed. Academic discharges were two-thirds and voluntary resignations were one-third. Dean Shields attributes the shift in the proportion of voluntary resignations at present to the fact that the men, let's say our young men of today, can't seem to take the military aspects of the academy, the regimentation, the indoctrination, and the confinement. Either they can't take it—or don't want to take it. Also there is much less of a need for free education, which was a large factor over 10 years ago. Of the academic discharges, he found that they have a low high school ranking, an average of 67% (this seems a little low to me, that is, that Admissions would accept applicants like that). We're talking about an average of 67% with low high school ranks as compared to 10% of those remaining at

the academy in that category; and when they fail they usually fail in Math. and the sciences.

Selection from a negative point of view is represented by our screening efforts in the Mental Hygiene Unit and by the screening carried out under our auspices in about 90 Military Examining Centers around the country and a few military bases overseas. I say negative, because our emphasis is more on screening out rather than selecting in. During the year preceding admission, slightly more than 5000 applicants are examined in these centers and by the personnel of the Mental Hygiene Unit of the Medical Department at the Naval Academy. The screening is accomplished via two instruments, a clinical interview and a Personal History Questionnaire. The Personal History Questionnaire was first developed in 1954 by then LT Allen McMichael and was a 20-page affair that was administered to the class of 1958. Since that time it has been pared down to seven pages and looks something like this. . . It invites a great deal of background history, such things as family information, a statement of the individual about himself, and such items as how he did in school, athletics, and so forth; extra-curricular activities, also what he expects he will find at the academy. Now on the back of this booklet there is a place for the clinician to put his rating of the candidate and make comments about his personality or what he might foresee as psychiatric or psychological difficulties. The rating system is a 5-point scale: 1, 2, 3, 3-, and 4. Point 1 is above average potential for making a success of the academy; 2 is average potential of making a success of the academy; 3 is below average; 3- is someone whom we think will not make it—we would not want him to be there or to come—but we can't justify a diagnosis; 4 is a psychiatric diagnosis, and the man is psychiatrically dis-

qualified. Now, what happens to the 3-'s; as a matter of fact, what happens to the ratings? These ratings are not seen by anyone except by us in the Mental Hygiene Unit, that is except for some of the 3-'s. If a man is medically disqualified on some other factor, eyes or weight or something like that which can be waived, his records are referred to a committee called the Academic Board. Our rating and our questionnaire go with these papers. Obviously, if they see that he is a 3-, they are more apt to reject him.

How good is this system? Insofar as is known, there has been only one study of this personal history questionnaire, done by Mark Goldstein or at least completed by him. It was started by several others in 1955. He found that this questionnaire could be broken down into 163 items that could be put on a computer. He couldn't computerize open-ended statements content-wise, but he was able to take the length of them, one sentence, three sentences or more. He found that these questionnaire indices could distinguish between the top 100 people in the class and a random sample of those who dropped out of that class, 74 out of 100 times. These findings were based on the Class of 1964. Out of these 163 items on the questionnaire, 18 appeared to have considerable weight. Using those 18 items alone, the top of the class could be differentiated from the bottom 70 out of 100 times. One of those 18 questionnaire items was the psychiatric rating. When he ran the 18 through the computer the importance of the psychiatric rating dropped off considerably. The psychiatrists' or psychologists' ratings alone could distinguish only 60% of the time in terms of class ranks. From these findings it may be argued that the questionnaire data alone, even though it was not meant to be a screening instrument, could do a better job than raters can do alone.



Notice the rough graphs, if you will. The abscissas represent the 1, 2, 3, 3-, and 4 psychiatric ratings, Graph A being the frequency distributor for the 5202 candidates for the Class of 1970 examined across the country in 70 different centers. It can be readily seen that Number 2 is the top category. These are the men who have an average chance of success at the academy. And 1's and 3's are pretty evenly distributed. Graph B depicts the frequency distribution for the 1339 men accepted for the Class of 1970. Note that there are a few more 1's and a few more 2's than found in Graph A. Graph C contains the men out of the 1970 class who were seen in the Mental Hygiene Unit and who are still at the academy. Slightly more 2's, more 3's and 3-'s are seen, but one should note the low frequency of 3-'s. Graphs D and E are those that were separated from the academy. D is based upon those interviewed by the Mental Hygiene Unit; those in Graph E were not seen by the Unit. Here again there isn't a great deal of difference between the two profiles. This finding raises some question as to the effectiveness of these ratings as screening techniques. Apparently the Admissions Office is really beating us to the punch in selecting and we are contributing very little from our field except for 4's and 3-'s. Hopefully some possible remedies for the situation may arise during this workshop.

One further thought—perhaps because of too general an initial conception of our task by our predecessors, we in the Mental Hygiene Unit and those in the outlying examining centers are attempting to predict a criterion which is too broad, causing us to go beyond the limits of our competency. Possibly we should be attempting to predict or identify those midshipmen who will leave the academy for reasons primarily of a psychiatric nature, for example, emotional difficulties of one kind or another. This leaves the remainder of the selection process quite properly to the Dean of Admissions.

SELECTION FOR SPECIAL ASSIGNMENTS: AQUANAUT SELECTION

Roland R. Radloff, PhD

Naval Medical Research Institute, Bethesda

As the title of my presentation indicates, it is based on the work I have done in connection with Project SEALAB. Before beginning, however, I'd like to enter a disclaimer. My research on SEALAB was not directed towards selection of aquanauts at the time that I did it. There have been questions raised regarding selection since, and I have been encouraged by SEALAB operating personnel to pursue research which will bear directly on selection for SEALAB III. Unfortunately, in addition to the fact that the research was not focused on selection, neither am I a specialist in selection. I am a social psychologist. Therefore, with these disclaimers, any unsophisticated gaucheries that I commit in discussing selection will, I hope, be forgiven.

I will concentrate on three major points here. First will be a brief discussion of the unique features of the SEALAB environment and the aquanauts themselves as they bear on selection. Second, I will talk about performance and adjustment criteria; and finally, about appropriate predictors of those criteria. It may seem as though I am putting the cart before the horse in discussing criteria before predictors, but I hope it will be apparent why I have chosen to do this.

I trust that most people in this room know a fair amount about the SEALAB Project and that I won't have to describe the environment in detail. An important feature of SEALAB relevant to selection is the small number of men involved. The entire group of aquanauts for SEALAB II numbered only 28 men. This stands in contrast with other groups discussed at this conference, Marines, Navy Enlisted Men, Naval Academy entrants, and so on. In considering selection for a very small group of

men versus mass selection one faces a different set of problems. SEALAB selection is more along the lines of selecting astronauts or men to climb Mt. Everest. I think the group that I found closest to SEALAB aquanauts are the men who winter-over in the Antarctic. Eric Gunderson and Paul Nelson have been working with selecting such men for quite a number of years.

Also important for selection are the peculiar features of this environment. It is an exotic environment. In SEALAB the participants breathed gas under one hundred pounds of pressure, with 85% helium content. It was a very demanding environment. While they were on the bottom for only fifteen days, this was quite a long period considering the high stresses they underwent. Also they were quite confined and their work demanded close coordination and close team work among the various teams. These features also distinguish SEALAB from environments in which Marines, Navy and Army enlisted men, and other groups discussed here have served.

So much for the environment. What were some of the salient characteristics of the men? Participation in the program required a high level of skill, particularly diving skill, but for other sub-specialties within diving as well. For example, the mean diving experience in this group of 28 men was eleven years. This experience factor, which was used in the selection of these men meant that they were much more experienced than say the entrants to the Naval Academy, Navy Enlisted Men, and so on. Thus they knew a great deal about the situation they were entering; also they are much older. The average age of this group was thirty-five. They were volunteers in the true sense in that they weren't escaping from the draft or accepting the lesser of two evils. They had to be chosen to get into this situation. They were highly motivated. I think high motivation characterizes groups such as this. Further, they were probably quite a bit better than the average of men in their specialty, as is the case with men volunteering for duty in the Antarctic. Looking at men who enter special environments in gen-

eral, I think they will be well above average in skill, ability and experience.

I think these characteristics of the environment and personnel have important implications for the criteria by which the performance and adjustment will be judged. I will turn now to the criteria.

Rather than gross criteria, such as pass-fail, retention or separation or advancement in rate, which are commonly employed when dealing with a large group of men, I think for a group such as SEALAB aquanauts one has to use multiple criteria, relatively fine grained, as fine as you can get it, to try to assess on as many dimensions as possible the relative performance and adjustment in this situation. The model that I used for SEALAB criteria derived from the work that Eric Gunderson and Paul Nelson have done in the Antarctic. They examined three conceptual areas of performance and adjustment. First, there is the task area—how well a man performs his job. Second, the social area—how well does he get along with the other men in the crew. Does he contribute to their performance and adjustment or does he detract from it? And finally, an area defined as emotional adjustment or emotional stability.

Let me preface this discussion of performance and adjustment in SEALAB with a phrase borrowed and adapted from Newell Berry. In talking about Marines, Newell says that there is no such thing as a poor Marine, there are only good, better, and best Marines. Similarly, there was no such thing as a poor aquanaut. In SEALAB the men were outstanding, without peer, and god-like.

Attempts at humor aside, a very high level of performance and adjustment were maintained in SEALAB, so the departure from the norm can be viewed in the upward direction.

On a conceptual level, task, social and emotional adjustment and performance are good starting points, but how in the world do you get at them? How do you measure these things? Down at the data level? This is what I'd like to concentrate on for the next few minutes.

In observing SEALAB aquanauts we had advantages which I think were without parallel in studying small groups under stress. We were able to see and hear these groups over closed circuit TV and audio channels, inadequate as they were. This is in distinct contrast to what Eric Gunderson faces in the Antarctic. He sends men off with a packet of questionnaires and hopes that they will fill them out during the time they are completely unavailable to him for six months. The only data he has is the reports that do come back. In SEALAB we were fortunate in having TV access which enabled systematic round-the-clock observation of all three teams. We gathered a variety of measures, and managed to tap each of the three areas—task, social, and emotional.

a good deal of kidding and ribbing of people who were always on the phone. Also it looked to us as though use of the phone reflected some negative association with the group. That is, if he wasn't a real core member of the group, maybe a diver was getting his social satisfactions outside by making phone calls. This is a good example of an unobtrusive measure which I think it is very important to have in a situation like this. The number of phone calls ranged from zero for several of the men to thirty-eight for the top number. This is quite a range to look at. It correlated very highly with a number of other face valid measures. For example, high and negative correlations were found with: the amount of time a man spent in the water; with sociometric choice; and with an objec-

RANK ORDER AND WEIGHTS OF CRITERIA FOR SEALAB II AQUANAUTS

Variable	Weight	Category	Type
1. Leader Rating	+ .75	T	Rating by Others
2. Phone Calls	— .74	S (e)	Official Record
3. Diving Time	+ .72	T	Official Record
4. Fear Scale	— .68	E	Self-Report
5. Post Peer Choice	+ .65	S	Rating by Others
6. Time in Lab Area	— .60	T	Observation
7. Gregariousness	+ .49	S	Observation
8. Meals Missed	— .49	T	Observation
9. Meal Prep. & Clean Up	+ .01	T-S	Observation
T = Task		S = Social	
		E = Emotional	

I have listed these variables in the rank order in which they contributed to weightings in a factor analysis. I have no great confidence in factor analysis as a way of telling you anything about your data except, I think in this instance it did organize the data and tell us which things were weighted more heavily in contributing to the assessment of the performance and adjustment in SEALAB. The measure which weighted most heavily (which is encouraging for those studying inaccessible groups in remote environments) was leader rating with a loading of + .75. The variable weighted next most heavily was outside phone calls. This may sound like a peculiar variable as many of the measures presented here might, so I'll discuss briefly the validity of this measure. We chose to examine this variable in detail because, through observation we could notice the telephone they had in SEALAB was used differentially by various divers and further, that there was

tive measure of gregariousness. I wish I could go into detail on all of these criteria because many of them may appear strange, but I think they represent the kind of thing which must be used as criteria for a group like this. That is, multiple criteria, looking at as many different facets of behavior in as great a variety of ways as possible. That's enough for phone calls. They weighted — .74. That is, the more phone calls a man made, the lower his score overall.

Diving time was right up there also, + .72. This was perhaps the best criterion of work performance. It was an official Navy record that was kept in SEALAB and criterion data such as this are, I think, without parallel. It is similar to the flight record if you are looking at performance of pilots.

The only self-report measure which was any good, and we had a number of them, was the mood check list. The fear scale on the

mood check list rated —.68 on our factor loading.

Post peer choice (another hint on methodology here, we used only positive choices) asking them "If you could go down again, which five men would you choose," loaded +.65.

Another variable, which I can't take time to discuss, was time in the laboratory area. We observed where men were in various parts of the capsule and this measure was the proportion of time each man was in this area. It is a task performance type variable and it's rated —.60.

The measure of gregariousness, based on 450 separate observations on each man, was the amount of time he was in the company of others, and how many other men were with him. This is a dust bowl empiricist with a vengeance definition of gregariousness. It rated +.49. Meals missed (we looked at every meal that was eaten, and recorded if a man ate with the group or not) weighted +.49. Meal preparation and clean up (this was assigned and each man was supposed to participate in equal amounts in this, but they, in fact, did not; many men participated more than others). Interestingly enough this did not weight at all with these other variables. It loaded by itself, +.01.

The important thing I would like to point out about this criterion variable is that by analyzing performance and adjustment this way you can get a factor score and can rank order each man.

There are four variables in the task area. They are: leader rating, diving time, time in the laboratory area, and meal preparation and clean up. Three variables are in the social area: the number of phone calls, post peer choice, and gregariousness. Two in the emotional area: number of meals missed, and self-rating of fear on our mood adjective check list.

I'd like to point out that there was a low reliance on self-report, that is, on asking a man to evaluate himself. With the exception of the fear scale, every variable depends upon some other method of measurement. Two of them depend on official records: the phone calls and diving times. Two were reports by

others: leader ratings and peer choice. And the last four were measured by objective observation.

I think that in looking at criterion measures, that is performance and adjustment criteria in groups like this, these are the kinds of measures that we have to aim towards. We did collect a great deal of self-report data, but the self-report data did not correlate with any of the objective data. It did not hang together meaningfully at all with any other rating of the men, with the exception of the mood check list. I'd like to discuss this in more detail, maybe in the discussion sections.

There is another important concept about looking at a criterion in this way—I don't think you can talk about absolute criteria in a group like this. Criteria will be relative. On all of the measures we used a Z-score, either for the man's own team or for the total group where it was appropriate. In other words, was he doing as well as, better than, or worse than the other men in the group on a particular measure.

Thus we feel that we have a good criteria. We can tell how well men perform in this environment, but what we really need are predictors. Unfortunately, we don't have much in the way of predictors and I'm not even going to go into what we do have. Rather, I'd like to discuss what we should have in the way of predictors.

First of all I think our predictors should be objective, they should concentrate on trying to measure past performance, social adjustment, and emotional adjustment in similar situations and should be selected on that basis. For example, if we are looking at divers, the best predictor, I think, is how much a man has dived previously. Unfortunately, on SEALAB we had only a gross estimate—the number of years each man had dived. There is a development here which I think is very significant for methodology. Lt. Tom Berghage at the Experimental Diving Unit has developed a method of recording dives for automatic data processing made by all divers in the fleet. This record will come into a central data bank and several years from now when a great number of men are needed

for a situation such as this, a complete history of a man's diving can be run off. This will be extremely valuable. It will be, I think, the best data that could be utilized. It will give a great deal of detailed information about a man's diving experience. It will be, I feel, one of the best predictors of performance in this situation.

The men on SEALAB II had eleven years of experience in diving. In subsequent groups they probably won't have that much experience, but they will have enough experience so that ratings from instructors and supervisors can be utilized. But I don't think we can count on standard fitness reports here because they are probably not too relevant to what we are looking at in this situation. In support of this contention I can cite a measure we used on SEALAB II. We had a measure of adult and juvenile misconduct which was slightly **negatively** correlated with performance in SEALAB. Maybe we had some very active and gung-ho types in SEALAB who act out. It may be that the best performers in this situation are not the ones who have completely clean records so far as civilian or military service are concerned. Such misconduct could be reflected in and invalidate normal fitness reports.

I am concentrating on gathering for SEALAB III ratings from instructors and supervisors. I realize that instructor rating forms are in bad repute, largely because of the problem of halo effect. If you have seen a recent article in the **American Psychologist** entitled "Mine Eyes Have Seen a Host of Angels," I think you have a good idea of what I am talking about. Even if you haven't seen this particular article, I am sure that if you have worked with leader rating forms you know what the problems are. I think one reason that rating forms are so poor is that there is probably not enough room at the top. So in developing these forms for SEALAB III I've tried to compensate for halo effect by forcing the rater to say this is the best man that I've ever known, I've only known one who is better than he on this characteristic, maybe two or three others, and so on. In the top 5% raters can make four distinctions. This is one approach that will be used. An-

other approach is to have the forms sent out by DSSP rather than by me as a researcher. Finally, the form will be sent to the man who has been closest to supervising the diver, rather than to the commanding officer of the vessel or the command to which he was attached. Ratings will be made by master divers and direct supervisors.

Other predictors that I think will give excellent payoff are peer ratings. They did very well on predicting the SEALAB II multiple criteria but this may have been due in part to shared method variance, since we did have peer ratings in the criterion itself. Also of use should be physiological predictors. Captain Bond has an ingenious idea of developing a film similar to the Lazarus film on emotional reactions, but using diving scenes and his comments on them to get physiological measures. Maybe we can talk about this later. I confess that I am quite naive in the area of physiological predictors but I think that they can be useful.

The philosophy guiding this approach to predictors is to utilize different types of predictors from those dependent on the man supplying information about himself. I think that when you have a volunteer situation like this where a man really wants to get into a program, that after he has said "I volunteer, I want to do this," he is giving you the last piece of valid subjective self-report information that he is going to give. After that point, everything is going to be structured to try to increase his chances of getting into the program. This comment does not, however apply to all self-report information. On SEALAB II information which correlated highly with the criterion was biographical information, objective information about a man's life history. I think there is a good reason why this should work out well, because biographical information is objective and is not subject to self-serving biases in its report.

In conclusion, the approach that I, as a non-selection specialist, propose is to try to get as great a variety of types of information about a man as possible, placing less reliance than psychology has in the past on verbal report, on a man's report about his own states

or about himself, both on the prediction and the criterion end. I propose the use of a conceptual scheme such that all aspects of a man's performance will be judged. Because all aspects of adjustment are crucial in exotic environments and in close-knit situations such as this.

It is clear that the various indices of adjustment used on SEALAB II are not orthogonal to each other. This replicates what Eric Gunderson and Paul Nelson have found in the Antarctic. Namely, that a man makes it in all ways or he doesn't make it at all. He works well, gets along well with the group, and is adjusted well emotionally. This is not necessarily so of men in normal life situations. A man can do his job well while developing an ulcer or fighting with his fellow workers. It is possible. I think that such adjustments are less possible in exotic environments.

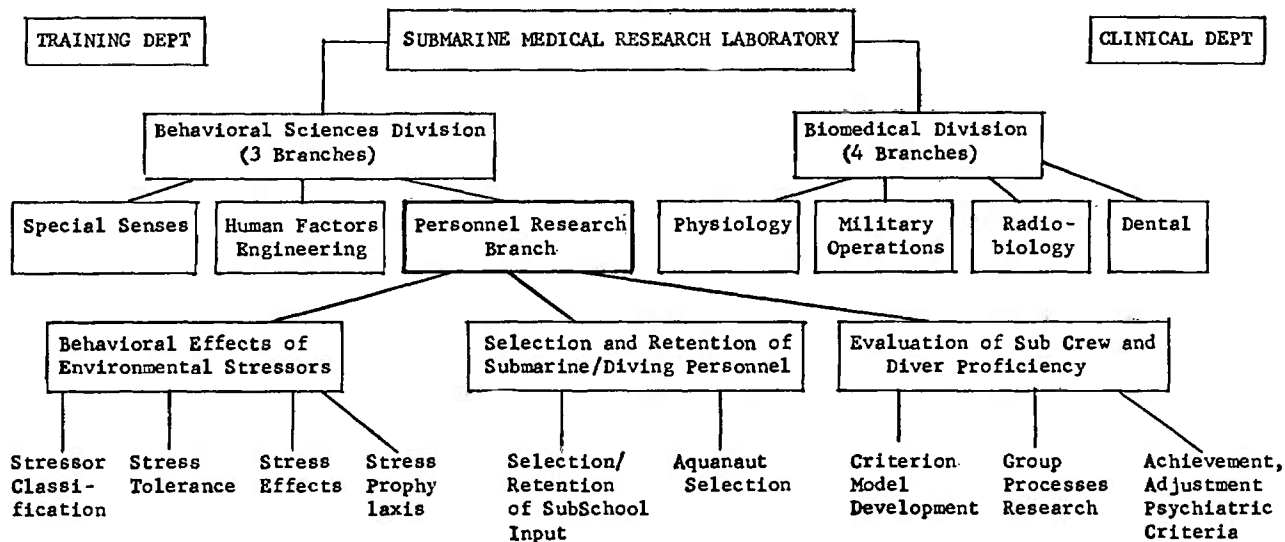
THE SUBMARINER SELECTION PROGRAM AT NEW LONDON

Benjamin B. Weybrew, PhD
USN Submarine Medical Center

In the preceding paper, Dr. Radloff mentioned that some of the staff members of the Personnel Branch of the Submarine Medical Research Laboratory have been involved in one way or another with SEALABS I and II. Though admittedly our contributions have been minimal in these programs, one goal of our Branch program has been and still is to collaborate in the development of efficient selection methodology for aquanauts and men for other high-risk duty. Our Submarine Escape Tank Training facility with a sizeable staff of experienced divers attached as instructors provides an excellent situation wherein various selection approaches may be subjected to preliminary field testing. In this connection a series of studies started in 1957 are still underway. But in the 15 minutes allocated to me (in consonance with the Workshop Title) I wish to focus only on the submariner selection program as it currently exists.

First, what is the position of the selection subtask in the total Branch program? The structural chart depicts this organization.

NAVAL SUBMARINE MEDICAL CENTER



It is seen in Figure 1 that the Personnel Research Branch is one of the three branches making up one of the two Laboratory Divisions, the Behavioral Sciences Division. Whereas collaborative studies involving inter-division and between-branch (within division) staff members (e.g., the hyperbaric helium-oxygen studies preliminary to SEA-LAB I in 1964) are undertaken, nevertheless most of the research effort is carried out by the small staff making up the Personnel Research Branch.

The Branch has three work units officially allocated to it (see abridged titles in the three bottom boxes in Figure 1). Indeed the content of the specific work units corresponds roughly to the three sections of the Branch. It should be relatively apparent I should think that the stated objectives of two of the three sections of our Branch smack squarely at the center of what one would suppose this Workshop program to be all about, i.e., the Personnel Assessment and Performance Evaluation Sections. Hopefully, as the Workshop sessions progress many aspects of these multifaceted problem areas will be examined in some detail.

Both civilians and men in uniform make up the Branch Staff. Incidentally, we have found that the uniformed psychologist and/or psychiatrist by-and-large is a very important member of the selection research team. This is particularly so in the Submarine Service wherein each submarine Commanding Officer (CO) occupies a rather autonomous role within the "subculture" delineated by the confines of the submarine. It appears that most CO's are more responsive to the medical officer and uniformed psychologist from the standpoint of granting permission for data of various kinds to be collected aboard his ship. Though our civilian staff do collect data in the field, we nonetheless have in the past relied heavily upon MSC and MC officers for initiating contacts, describing and indeed selling programs, direct supervision of data collection and the like.

Now, for a few specific comments about the manpower "pool" numbering 4500-6000 enlisted men and 400-600 officers annually. At this time a critical point needs to be made. With forty-one Fleet Ballistic Missile submarines commissioned, with each ship manned by two crews of 125-130 officers and men, with 40-50 additional, single-crew nuclear submarines—these facts coupled with a 40-50% or less over-the-board first reenlistment rate, the demand for highly trained nuclear submariners is indeed high. Some of the personnel managers apparently assume that the numbers and mixes of specialties needed to maintain an effective submarine force (both officers and enlisted men) approaches the number volunteering for this branch of the naval service. Presumably the BuPers quota system is based upon continuous (or at least periodic) inputs regarding recruiting practices, reenlistment rates, fleet personnel requirements and the like.

While varying from month to month, in order to maintain an effective Submarine Service a selection ratio approaching 90% is needed. Now, if you recall the Taylor-Russell Tables* the effectiveness of a selection program involving a test or tests of given validity is inversely proportional to the selection ratio imposed as well as the prevailing unselected success ratio. Therefore with selection ratios of necessity being as high as ours, our program may be properly labeled personnel screening, our test scores and observational techniques being a "sieve" designed to identify poor adjustment risks for the Submarine Service.

Before I make a few comments indicating the nature of our very modest attempts at developing a formal screening program in this context, let me make a few remarks regarding what we call "systems screening" to refer to the screening function afforded by the system itself, in our case, the training situation. Accordingly, the attrition rate for enlisted men at the various stages of the program, though varying considerably from time to time is as follows: At the basic Submarine School level 15-25%, at the submariner qualification level (approximately one year following Submarine School graduation) 4-10% and

*Taylor, H. C. & Russell, J. R. The relationship of validity coefficients to practical effectiveness of tests in selection. *J. Appl. Psychol.*, 1939, 23, 565-578.

at "shipping-over" time 40-50%. As one might guess with our selection ratio as high as it is the attrition of 15-25% early in the training period would seem to constitute an excellent screening device in and of itself. Moreover, Brogden's "dollar criterion" at least for enlisted men appears to be relatively favorable, since the cost of "failing" a man is comparatively little.* Hopefully some of the Workshop discussions will be directed toward an examination of the part "systems screening" plays in certain selection situations.

Though the selection "picture" may in the future change in the direction of a reduction in the essential selection ratio, we are, as I mentioned previously, presently engaged in psychological screening rather than selection. Stated another way, our program is designed to identify out of the volunteer pool those submariner candidates whose adjustment potential for the Submarine Service is minimal, keeping in mind that with the imposed selection ratio being as it is, it is much more desirable to make the mistake of accepting the marginal man who may not adequately adjust to the service (the Type II error for the statistician) than to reject a badly-needed "good" risk who would have in fact become a career submariner (Type I error).

Let me turn quickly to a brief overview of our screening (vice selection) program. Although the criticality of the submarine officer screening and retention problems may be as severe as that of the enlisted man, nonetheless my remarks today will pertain to enlisted men only.

First, consistent with some of Dr. Plag's findings reported a few moments ago, there are several population or demographic variables which provide a basis for identification of poor risks if our criterion is the primary one of basic Submarine School attrition, (keeping in mind our success ratio lying between 4% and 40% varies directly with the selection ratio imposed by the BUPERS quotas). Two of these variables are education level and age. For example, the success ratio

(SR) for high school (HS) graduates without controls on any other variables operating in the volunteer sample is of the order of 85% as compared to 50% for non-HS graduates, or, if the HS dropout is young, say less than 20, the SR shrinks another 8-10 percentage points. The selection experts in this Workshop group will immediately recognize some confounding involved here in part, at least, resulting from the well-known positive and quite "strong" relationship between ability and formal educational achievement. Accordingly, for the HS graduate above mean GCT, ARI, MECH (or any combination of the three) the SR at the Basic Submarine School level characteristically falls in the 90-95% range. Rated men particularly in the so-called critical electronics ratings enjoy virtually 100% probability to "get by" the Submarine School graduate-drop criterion. In short, if age, education and aptitude are taken into consideration usefully high reliability of our predictions can be achieved with respect to Submarine School graduation. Parenthetically, the predictive capability for these variables disappears as the criteria become more remote. Some of these remote criteria for which we have few if any reliable predictors are: qualify versus fail to qualify for submariners subsequent to Submarine School graduation, individual differences in underway performance, or ultimately, reenlist versus fail to reenlist (even if eligible).

As for our psychological screening program several general statements can be made.* First, empirically-keyed objective group tests by-and-large have proven more useful in our program than have group-administered projective-type tests. Second,

*There are three publications which present a brief history of submariner selection at New London as well as cursory description of the program as it developed over the past 15 years. These are:

Shilling, C. W. & Kohl, Jessie W. History of Submarine Medicine in World War II. USN MedResLab Report 112, May 1947.

Weybrew, B. B. Some general comments concerning personnel selection for hazardous duty —after 6 years of research. USN MedResLab Memo Rpt 62-5, Oct 1962.

Weybrew, B. B. Psychological problems of prolonged periods of marine submergence. In Burns, Chambers, & Hendler (Edits), *Unusual Environments and Human Behavior*. Glencoe, Ill: Free Press, 1963.

*Brogden, H. E. When testing pays off. *Personnel Psychol.*, 1949, 2, pp. 171-183.

the brief screening interview is not only impractical with our small professional staff (in the context of a 400-500 man monthly input) but, with test data partialled out, does not appear to be a useful screening technique in and of itself. Rather our approach has been to identify the "marginals" by means of paper-and-pencil tests then recall these men for a more detailed diagnostic interview and in some cases, further diagnostic testing. Integration of the test and interview data provide the basis for the final accept-reject decision. As a rule, approximately 10% of the enlisted input are identified by the tests proper as "marginals" to be subjected to more intensive study and evaluation—from 5-8% are rejected at this level.

A variety of tests and observational techniques have been used as selection variables, group TAT and group Rorschach, MMPI, psychomotor apparatus tests and so on. We now focus on three classes of traits presumably relevant for optimal adjustment to the conditions existing during prolonged submergence. These classes of traits are listed below together with a descriptive statement regarding the kind of paper-and-pencil test used to measure them.

1. Specialized Abilities (Aptitudes). The BuPers Basic Test Battery scores, readily accessible in each man's service record, are very useful for identifying poor risks, this is particularly so for standard score combinations of ARI, MECH and GCT.

2. Biographical Information. A machine scorable biographical inventory has been "custom-tailored" for the submariner candidate population, one form for officers, another only slightly different, for enlisted men. This fifty-item form is designed to obtain socioeconomic, sociological and demographic information to be used in one way or another to support decisions regarding rejection for the Submarine Service.

3. Motivation for the Submarine Service.

Several modifications of a paper-and-pencil questionnaire have been in the screening battery over the past 5 years. A multi-category response format extending from "Not at all like me" to "Exactly like me" is applied to each of 50 items designed to "tap" the most relevant goals or satisfiers (or the needs underpinning them). Using aptitude scores as a moderator variable, several keys have been constructed. Several of these scores provide useful information regarding performance and adjustment deficiencies turning up in certain men at the Submarine School level and after. This inventory is called the Self-reported Motivation Questionnaire (SMQ) and is described in two publications.*

4. Neurotic Symptomatology.

Utilizing the same multicategory response format as used for the SMQ, the present modification of this questionnaire contains 100 items, most of them of the MMPI variety. Several kinds of itemetric analyses have been done on several "batches" of experimental items, Wherry-Winer factor analysis, a Loevinger-type reiterative analysis and of course the usual internal consistency analysis. The most useful key derived from these items helps a great deal to identify the rather rare submariner candidate with neurotic traits that are potentially debilitating or at least handicapping, those prone to acute anxiety reactions, diffuse phobias and the like.

This test called the Personal Inventory Barometer (PIB)* has been used in several Submarine Medical Officer Qualification thesis involving the identification of groups of submariners within a given crew showing different levels of adjustment to submerged conditions.** An example of this type of a study may be found in the reference footnote.

*Weybrew, B. B. & Youniss, R. Personal Inventory Barometer: I. Development of the questionnaire. USN MedResLab Report 290, Aug 1957.

**Nigro, S. A. Psychological correlates of prolonged confinement in a closed atmosphere. SMO Qualification thesis. March 1964.

***OPNAVIST 5510.83
BUPERSINST 5510.11A, and
NAVMED P-5090, Personnel Reliability Manual.

*Weybrew, B. B. & Molish, H. B. Approaches to the study of motivation of officer candidates for Submarine Service. USN MedResLab Report 321, Oct 1962.

Rubin, Barbara & Parker, J. W. The Self-Reported Motivational Questionnaire. USN MedResLab Report 348, Feb 1961.

5. Attitudes toward the Navy.

In the early sixties, a series of instructions originating from OPNAV, BUPERS and BUMED*** laid the ground work for what has come to be called the Reliability Program, a variation of which is found in each of the services. Briefly and staying away from certain security aspects of the instructions, SMRL was instructed to conduct an initial evaluation and thereafter periodic re-evaluations of the general adjustment status of each enlisted man and officer occupying billets deemed critical in that the incumbent would have some degree of contact with nuclear weaponry. Without any elaboration, I am sure, even if you've never heard of the Reliability Program, you can guess the essential intent of the instructions. Although the instructions indicated that a detailed psychiatric examination was called for, the personnel input of 4000-6000 annually, taken in the context of a scarcity of qualified personnel to conduct such a procedure, made this approach impossible. As screening experts, what approach would you take toward meeting these requirements? Again without breaching security, you can guess we were being asked to identify the impulsive, the debilitated neurotic, the incipient psychotic particularly those with well-established delusional and/or chronic depressive symptomatology. As a start, we made an assumption that a valuable addition to our existing screening battery would be realized if a meaningful measure of individual differences in attitudes towards the Navy, towards the deterrence concepts and towards nuclear armamentarium, could be constructed. Accordingly, an experimental attitude questionnaire, the Personal Attitude Questionnaire (PAQ) was constructed. Application of the latest itemetric techniques to circa 175 attitude items resulted in an attitude scale with about 50 items. The response format was a vertical Likert-type scale (Figure 2), the content of the 50 or so items composing the final keys having to do with attitudes toward war and peace; e.g., "The FBM submarines are a threat to peace,"

"Better Red than dead" etc. While in a few instances the summed-scores from the PAQ as well as individual item responses suggest possibilities as a technique of identifying certain incipient pathology, we have not as yet acquired any firm validation data for any of the keys for any purpose. While it gives us little consolation, the information we have indicates the other services are having some considerable difficulty in implementing similar programs peculiar to their needs. Hopefully, some new ideas will arise during the Workshop sessions in the next few days.

How do these screening data interact? Using a multiple-cutoff in preference to a linear regression model, we have some contingency data for the BTB, PIB and SMQ scores with respect to the proximal criterion of Submarine School attrition. Table I shows these interrelationships for a sizeable sample of enlisted men.*

Table I

The Interaction Effect of Measures of Psychiatric Status, Motivation, and Aptitude upon Attrition in Enlisted Submarine School (N=1249)			
Variable	Unfavorable psychiatric status ^a	Favorable psychiatric status	
	Attrition (% of Frequency in Each Cell)		
High ^a Motivation	12.8 (N=109) ^b	4.1 (N=195)	
High ^a Aptitude			
Low Motivation	10.4 (N=192)	3.3 (N=121)	
High Motivation	16.7 (N=132)	10.3 (N=184)	
Low Aptitude			
Low Motivation	22.3 (N=193)	15.4 (N=123)	

*All dichotomies are at the approximate median of the test score distributions.

^bTotal frequency in each cell.

It is immediately seen that if the selection ratio could be of the order of 40% the three test scores (Aptitude, BTB scores, PIB & SMQ) alone would result in a success ratio of about 96% at this level of training. As I mentioned earlier, but I think bears repeating, partly resulting from variance shrinkage, the predictive validity of these tests (and I might add a variety of other tests and measures we have tried from time to time over the past decade) do not show useful validity in relation to more ultimate criteria of adjustment to prolonged submerged, or the all important index, sustained favorable

*Taken from: Rubin, Barbara & Parker, J. W. The Self-Reported Motivational Questionnaire. USN MedResLab Report 348, Feb 1961.

orientation for a full-retirement career in the service. I am sure others have similar validation findings to report, possibly in slightly different settings.

In summary, the Submariner Selection Program, while presently functioning somewhat effectively as a screening procedure, is hopefully flexible enough to operate reason-

ably well in a more specific selection role should the "supply/demand" situation for "high caliber" submariner candidates become more favorable. Without question, some of the ideas and approaches brought to light during these Workshop Sessions will be of use in the solution of some of the problems I have raised.

MRL QUESTIONNAIRE

NAME _____ DATE _____

RATE OR RANK _____ AGE _____ CLASS NO. _____

Education: Did not complete High School, Completed High School,
Some College, College Graduate (circle one)

		1	2	3	4	5
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		6	7	8	9	10
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		11	12	13	14	15
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		16	17	18	19	20
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		21	22	23	24	25
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		26	27	28	29	30
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		31	32	33	34	35
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		36	37	38	39	40
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		41	42	43	44	45
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		46	47	48	49	50
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

+4	_____	
	_____	Agree strongly
+3	_____	
	_____	Agree considerably
+2	_____	
	_____	Agree slightly
+1	_____	
0	_____	Neither agree nor disagree
-1	_____	
	_____	Disagree slightly
-2	_____	
	_____	Disagree considerably
-3	_____	
	_____	Disagree strongly
-4	_____	

NSMC Form 32

FIGURE 2

+4		
	Agree strongly	
+3		
	Agree considerably	
+2		
	Agree slightly	
+1		
0	Neither agree nor disagree	
-1		
	Disagree slightly	
-2		
	Disagree considerably	
-3		
	Disagree strongly	
-4		

51	52	53	54	55
56	57	58	59	60
61	62	63	64	65
66	67	68	69	70
71	72	73	74	75
76	77	78	79	80
81	82	83	84	85
86	87	88	89	90
91	92	93	94	95
96	97	98	99	100

SESSION II

CRITERION DEVELOPMENT

SESSION OVERVIEW

THE CRITERION PROBLEM IN PERSONNEL SELECTION

Randall M. Chambers, Ph.D.
Naval Air Development Center
Johnsville, Pennsylvania

The multifaceted nature of the existing personnel selection programs within the Navy as well as the multifaceted nature of the many missions, tasks, and job requirements within the Navy, results in the problem of criterion development being one of the most difficult yet most important problems in the entire field of personnel selection. Criterion development is an essential and complex aspect of the criterion problem, and the ultimate success and evaluation of any personnel selection technique, method, or program is highly dependent upon the criterion involved. In choosing from a group of men those most likely to succeed at the highest possible level of proficiency, there are usually a wide variety of immediate as well as long-range criteria, some of which are available and some which must be developed. In this first workshop session, as I see it, our objective is to review and discuss the various aspects of the criterion problem as it relates to other aspects of personnel selection methodology, and to formulate some recommendations regarding some of the key issues that hopefully will emerge from the session.

An acceptable criterion of success is crucial because it constitutes the basis for validation, for prediction, and for establishing reliability estimates. A criterion may be expressed in terms of a single characteristic, or in terms of a set of measures, as in a multiple regression problem situation. However, the criterion attributes, or sets of single characteristics, are usually combined to provide a composite measure of proficiency or often success/failure in terms of which prediction, validation, and reliability statements may eventually be made. Also, the

selection process itself is frequently a sequential one, the criterion at an intermediate stage for example may be considered a predictor for a later stage, and finally, the last measure of success or performance is often designated as the "ultimate" criterion. Thus, the criterion problem consists, in time, of a developmental process, passing through stages of observation, use, and evaluation involving prediction, validation, reliability estimation and re-estimation.

Criteria for personnel selection are identified and developed in a very dynamic way over time and in a variety of situations. This process includes the identification and utilization of a variety of criteria at early stages, and possibly some ultimate criterion at a later stage. This is especially true in selecting personnel for jobs in unusual or stressful environments and involving extremely difficult and hazardous tasks, or for certain complex man-machine weapon systems which have varying and changing mission requirements. It is here that the criterion development problem becomes quite difficult and complicated. Since human performance sometimes appears to be continuously modified by certain types of external events, predictions of future performance are usually made conditional upon the occurrence of a specific set of conditions. Thus the performance capabilities, interests, and motivations of man, as well as the characteristics of machines and environments, together with anticipated mission and task requirements, must be given appropriate weighting and consideration as a part of the criterion development problem. Temporal factors and relationships among human capabilities, machine capabilities, and mission requirements

must be included also.

The criterion should provide several other specific attributes. It should provide, for example, (1) the standard in terms of which the relevant predictor variables can be isolated, (2) the efficient and appropriate testing and evaluative procedures separated from the inefficient and inappropriate ones, (3) the relative weights determined for use in predicting future performance and in combining sets of observations or measures, and (4) maximum utility for maintaining appropriate mission effectiveness throughout the conduct of the selection program, and (5) an adequate range or depth of performance evaluative capability within each criterion attribute. It is not easy to meet the five criterion attributes listed above. Our technology approaches them, but it is not adequate. Much research and technical development effort is required with regard to each of them.

Also, the criterion must provide an adequate definition of the success continuum for the task or activity in question. The measure of success is sometimes assumed to lie along a single continuum, even though success is not unitary. Success is usually the result of a large number of separate abilities, skills, and personality characteristics. We should express the success continuum as a multi-dimensional variable with each dimension an independent component. It is possible, and is to be recommended, that the definition of the success measures or criteria have the following characteristics: (a) a single overall evaluation, (b) a weighted composite of the separately measured components, (c) a pattern index of these several variables, (d) a composite profile of the human and the situational requirements, and (e) a composite profile of the human (man) and equipment (machine) components and their relative utilities within specific situational requirements and mission requirements.

The merits of the criterion are judged to an extent by their (a) validity, (b) reliability, (c) discrimination or selectivity, (d) utility, and (e) availability. The validity of sets of criterion measures are evaluated in

terms of statistical evidence of intercorrelations among the measures. Face validity is also important, however. The reliability and discrimination of interview measures are evaluated in terms of indices of consistency over periods of time and over samples of situations. Utility usually relates to cost effectiveness, and availability usually relates to convenience of the measure and to the practical question of obtaining the criterion information within time and organizational constraints. Finally, it is essential that the criterion meet minimal standards for accuracy throughout all stages of development. The demands for precise and accurate presentation of the criterion data are sometimes difficult to meet, and frequently research is needed in order to determine the data requirements and to obtain the criterion data itself.

One of the problems which always confronts the personnel selection specialist is the source of data to be used as criteria. In order to facilitate later discussion, it may be helpful to list some of the sources which are frequently used. They are: (1) intelligence and aptitude test results, (2) proficiency and ability measures, (3) personality, interest, and motivation tests, (4) biographical and autobiographical data, (5) interview data, (6) adjustment indices, (7) supervisory and self ratings, (8) results of factor analysis studies, (9) performance on tasks and job components, (10) performance in simulations, (11) performance in field and operational situations, (12) results of on-the-job training, (13) sustained performance on difficult tasks, (14) critical incidents, errors, mistakes, etc., (15) self-selection data (e.g. volunteering), (16) biomedical indices, (17) hereditary and family background, (18) health records, fitness reports, age, (19) socio-economic factors and history, (20) educational background, and (21) special aptitudes, abilities, and motives.

This list is not intended to be complete, although it does give the sources which are frequently used. Whatever the source of data, however, for use in the criterion development program, it is important that it be related to the job requirements, e.g., the

personnel requirements for the job, and more specifically, the situation within which the job must be performed. Sometimes this includes the specification of unusual environmental conditions, stressful missions, or utilization of complex equipment within manned weapons systems. It is important that specifications for these special conditions, situations, and equipment be considered as an essential part of the criterion development process. The statistical problems in accomplishing this can become overwhelming, however, and a great deal of judgment and careful evaluation are necessary as one proceeds through the process. Some of these methodological problems involve rules for combining measures, for weighing measures, predictor variables (both dependent and independent), time-dependent variables, and overall system performance requirements.

At this point it may be well to summarize the major types of criteria which are commonly used in selection programs. They are as follows: (1) simple (task component) criteria, as related to a specific task or job, (2) complex (total task) criteria, as related to a specific task or job, (3) composite criteria, consisting of weighted components, (4) global (holistic) total criteria, consisting of performance, personality, motivation, and socio-economic factors, (5) self-selection criteria, (6) immediate criteria, (7) long range criteria, (8) individual criteria, (9) small group (team) criteria, (10) man-machine criteria, (11) situational criteria, (12) statistical (predictive) criteria, (13) mission success criteria, and (14) pattern (profile) criteria.

My main personal interest at the present time with respect to selection and specifically the criterion problem has to do with weapons systems and the problems involved in developing, testing, and manning weapons systems. This involves at least the following considerations: (1) the human factors, characteristics, and skills required in operating these weapons systems effectively, (2) the specific requirements of the weapons systems, (3) the environmental hazards, stresses, and requirements, (4) the task re-

quirements, and (5) the total mission requirements. The design of weapons systems is to some extent dependent on human factors. By the same token, the needed personnel requirements are to some extent dependent on the weapons systems requirements. There are man-machine trade-off requirements as well as necessities for allocating man-machine functions within any given system. Further, the man may be required to function as a part of a system with a team, or with a machine, or both. Finally, he may be a part of the loop for mission success, and his skills and abilities may vary according to the criticality of his performance for total systems effectiveness as well as the conditions under which he must work. Our interest, therefore, in personnel selection is to a great degree related to the manning requirements of current and planned weapons systems within the Navy. Some discussion of the criterion problem as related to weapons systems development is therefore required. Hopefully, some attention will be given during this Workshop Session to problems of manning systems functioning in unusual and stressful environments.

Some of the projects and weapons systems on which we in the Psychology Division at NADC are currently working are summarized in Table I. These are funded on-going projects for which major effort is being expended. Some of them are in the far future and some are current.

TABLE I
SYSTEMS AND PROJECTS RECEIVING MAJOR
HUMAN FACTORS SUPPORT

ASW A-NEW	Satellites & Spacecraft
ASW Patrol A/C	Weightless Simulation
AEW/C	Flashblindness Protective Devices
E2B	Displays
F 111B/Phoenix	Retinal Camera
720B/707	Controls
Fighter Attack	Vestibular Truck
F4 Spin	Anti-blackout equipment
Cockpit Instrumentation	G-Protective Devices

TABLE II
HUMAN FACTORS CATEGORIES

Sensory Thresholds	Computational Skills
Visual Performance	Team Performance
Auditory Performance	Endurance
Orientation Mechanisms	Fatigue
Motor Skill Performance	Psychophysical Measures
Piloting Performance	Performance Measures
Other Operator Skills	Simulation Technology
Learning & Conditioning	Test & Evaluation
Judgment & Decision Making	Experimental Design
Attention & Vigilance	Simulation
Immediate Memory	

on the figure indicate which category is necessary for each system. For example, the ASW-A-NEW system requires active support throughout all 21 categories. The other systems, for example, each require effort in most of these human factors categories. It should be pointed out that these are not as compartmentalized as they appear since the interactions among the various categories and systems could not be diagrammed, and since time-in-depth aspects could not be diagrammed in this figure. By having different

TABLE III

	Vision	Hearing	Orientation	Target Detection	Target Recognition	Sensory Thresholds	Simple Motor Skills	Piloting Skills	Operator Skills	Judgment	Attention	Memory	Target Identification	Endurance	Individual Differences	Team Performance	Experimental Design	Psychophysical Measures	Performance Measures	Simulation and Testing
ASW A-NEW	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ASW Patrol A/C	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
AEW/C System	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
F 111 B/Phoenix	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Fighter Attack	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
720B/707 Turbulence	X		X			X	X		X	X	X		X	X		X		X	X	
32 A(MOL)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
F4 Spin	X		X				X	X	X	X	X	X		X	X	X		X	X	
Cockpit Instrumentation	X	X	X	X		X	X	X	X	X	X			X	X	X	X	X	X	X
Satellites	X		X			X		X	X	X	X		X	X	X	X	X	X	X	X
Vestibular Truck	X	X	X		X											X	X	X	X	
Retinal Camera	X															X	X		X	
Eye Protection	X															X	X	X	X	
G Protection	X					X	X	X								X	X	X	X	

Table II summarizes some of the primary human factors categories in which the Psychology Division has active research and human engineering studies in progress.

In Table III an attempt has been made to show a comparison in terms of the various human factors which are involved in the various systems. This table suggests some of the personnel characteristics required to operate each system. This table attempts to show which human factors categories tend to be required repeatedly over the various weapons systems. The x's within each square

persons specialize in different category areas, however, some attempt is made to economize in man-power requirements by this procedure. It is based on the premise that personnel having these specialized skills will be available to operate each system.

In human factors research and human engineering, the Psychology Division proceeds on the assumption that the capacities and abilities of the human operator are set within certain natural limits (with a certain variation due to training and individual differences), and that one way to design good

weapons systems is to adapt the systems properly to human capabilities. In collaboration with engineers and scientists, the human factors psychologists at the Naval Air Development Center attempt to develop new and improved equipment concepts and man-machine systems which will simplify the operator's task and improve the probability of mission success. The successful design of equipment for human use requires consideration of the man's basic characteristics. Among the major ones are his sensory capacities, his muscular strength and coordination, the speed and accuracy of his motor movements, his body dimensions, his perception and judgment, his native skills, his capacity for learning new skills, his optimum work load, his basic requirements for comfort and safety, his ability to tolerate environmental stress, and his ability to perform skillfully in a reliable and consistent fashion.

In the most specialized sense, the human operator and his machine are regarded as a single entity—the **man-machine system**. This concept, perhaps more than any other, now complicates the problem of attaining useful selection criteria more than any other concept which has been introduced to selection methodology during recent years.

The man-machine system is considered to be an assemblage of human operators and machines which are in significant communication with one another and which are performing tasks sufficiently well defined so that independent and dependent variables and criteria of systems performance may be operationally specified. The performance of the man and the machine are both involved as interaction processes. Implicit in the concept is the functional allocation of man-machine components, and the assurance of optimum man-machine interface conditions. Background data and research in support of these concepts always include the experimental analysis of operator function. For any given man-machine system, the human factors specialist involved in the design of the weapon system must carry out his own experiments to determine the important function variables of man and of the physical

system in order to make required trade-offs of task components within the total system. Methodology holds an important place in the human factors field, since the human characteristics must be matched with the machine at the intellectual as well as the motor skill level of operator function. Appropriate matching of machine and man characteristics is essential for effective system performance.

Research and development programs in human factors originate from a variety of sources, and the Psychology Division attempts to meet the requirements of these programs by assigning specific persons the major responsibility for conducting the program and by calling on other personnel or units for assistance when needed. These requirements must be considered at all stages and levels of criterion development. Also, these requirements tend to influence the degree and types of personnel performance capability which will be designed into the particular weapon system during design and utilization phases.

In view of the varied sources of assignments, personnel have different degrees of responsibility in the various programs. They may be entirely responsible for the direction of the program, or they may be responsible only for the human factors aspects. The Plans and Programs Office at the Naval Air Development Center, and the Director, AMRD, frequently make the determination of the degree to which individuals and/or groups will be given responsibility for the accomplishments of these assignments. Frequently a psychologist within the Division may be designated as either "Project Officer," or "Principal Investigator," or "Principal Administrator," or "Project Engineer." These assignments are on a project assignment basis and do not affect the internal organizational structures within the Psychology Division.

When a program is received or initiated, it is examined from at least two points of view. One point of view involves the effects of the weapon system on the aircrew, e.g., the effects of the environment produced by the weapon upon the aircrew members. Included in this are the life support aspects and the

protection of the crew against hostile environmental stresses such as acceleration, temperature, atmospheric constituents, noise and vibration. Information concerning the human factors involved in the interaction of life support and weapons systems is obtained by research and testing, using simulation environments, or real stresses, to study the effects of protective systems. A large portion of this effort is in terms of the psychological performance and physiological effects of acceleration stress on crew members. For this work for example, the human factors personnel have available the world's largest and best instrumented human centrifuge. The effects of the system on the crew are studied in this device. This centrifuge is used principally for simulating aerospace vehicles and for evaluating human performance capabilities and limitations during exposure to the acceleration environment typical of certain types of missions in certain types of aerospace vehicles. Some of the airplanes simulated to date include: the 720B, the A2F, the A3J, and the A4D. Some of the spacecraft which have been simulated for studying effects of crew performance are: X-15, X-20, Mercury, Gemini, and Apollo.

Other categories of effort in human factors include command and control. Under command and control, the interaction of the operators of the weapons systems with the appropriate equipment are studied. It is in this area that the Psychology Division is currently providing personnel to work in support of several weapons systems, such as A-NEW, AEW/C, E2B, ASW Patrol A/C, and F-111B/Phoenix. In these, a simulation is a major portion of the effort (see Figure 2 for a schematic diagram of some aspects of these procedures). In such weapon systems, psychologists are usually responsible for determining operator requirements through the conduct of interviews and such procedures as timeline analyses and information flow analyses. They also direct simulation programs which have as their goal the development of proper interfaces between man and equipment. Such simulations also provide preliminary evaluations of system effectiveness.

The simulation efforts vary in complexity.

They usually start with relatively simple mock-ups which allocate work spaces and equipment organization, often including display and control systems. The more involved phases of simulation usually include dynamic activation of the displays and controls operated via a general purpose digital computer. In the more complicated problem areas motion simulation is included. In some phases of weapon systems operator force fields have significant importance. The Psychology Division uses fixed-base simulation techniques in attempting to obtain realistic methodologies for performance measurement. On some occasions simulation includes in-flight evaluation of operator performance. The results of human factors evaluations conducted at NADC are used to improve the efficacy of the systems. Throughout this entire series of simulation, attempts are made to obtain new basic data while the human engineering is conducted on the weapons system simulations. The laboratory research is attempted in conjunction with the weapons system development. The ultimate goal is to design and develop better weapons systems through human engineering and human factors research with appropriate consideration of personnel requirements information and availability of personnel who can operate the system effectively within the range of expected environmental stresses. The criterion problem is not solved, but some suggestions regarding its ultimate nature are considered.

In this section of the paper I have attempted to outline some of the major human factors categories which are currently being studied at NADC, and some of the weapons system projects to which these human factors categories have application. Some of the ways in which these human factors specialists work to support weapons systems development are outlined. An attempt is made to show some of the ways in which the Psychology Division currently studies the interrelationships among the various systems and factors. Groups of human factors engineering categories are reviewed in terms of their current contributions to weapons systems development. Some categories of human factors contribute to a relatively large number of

weapons system problems. Similarly, certain weapons systems contribute to the accumulation of new human factors data, and this data is found to be of use in the design and development of man-machine systems. This has important implications for selection criterion development problems. Recent attempts to improve methodology for measuring performance and methodology for optimization of function allocation are indicated. Simulation is emphasized as a primary research approach. Effects of weapons systems on crew performance, as well as effects of crews on system performance, are emphasized.

Needs for additional personnel in human factors are suggested. Technical needs for improved human factors research and personnel skill selection problems include: (1) improved methodology for crew performance measurement; (2) improved methods for optimization of function allocation among men and machines; (3) improved simulation techniques; (4) new measures of operator performance; (5) additional basic data on human operator performance capabilities and limitations within specific weapons systems; (6) additional information on the interaction effects of **combined** variables on operator performance; (7) the need for basic psychophysiological methods for relating physiological function to crew performance; (8) a means for identifying human factors requirements for any given weapon system early in system development phases; (9) a means for insuring that human factors support of weapons systems development is started early during the inception phase of a system, rather than after the system has been manufactured and been in use by the fleet; (10) a more efficient means for obtaining systems information from weapons systems project groups, and (11) realization of the fact that the research previously conducted does not necessarily have application to a new weapons system concept and consequently, that there is frequently a need for obtaining new human factors data which may be specifically relevant to any new weapons systems concept under consideration as well as to available or expected requirements of existing systems.

It is hoped that during the discussion period to follow, some of the issues may be resolved.

At this point a return to the more traditional aspects of personnel selection criteria problems should be attempted. Some of the concepts outlined above are somewhat new, controversial, and related to specific situations of somewhat limited scope. The following outline contains some of the guidelines for personnel selection, classification, and criterion development within the Navy!

A. **GOAL:** Selection of personnel who will be physically, mentally, and operationally fit to perform the tasks for which selected.

B. **SCOPE:** Conservation and proper utilization of manpower is essential where the ever changing role of man becomes more important in our weapons systems of the future. We must have the right man for the right job, at the right time.

C. **REQUIREMENTS OR SUB-GOALS:**

1. **Physical**

a. Establish minimum and maximum physical standards for tasks needed in the future through categorizing tasks where there is carry over from one task to the other.

b. Studies aimed at early detection of disease in all systems of the body. Example: cardiovascular disease.

2. **Mental**

a. Establish mental and emotional standard for the task.

b. New techniques for predicting and categorizing:

(1) Motivation

(2) Stress tolerance

(a) Boredom

(b) Sleep deprivation

(c) Isolation

(d) Acceleration

(3) Personality type and relationship to performance

(4) Psychiatric stability

(5) Mental capacity (IQ)

(a) Working alone

(b) In groups (team)

(6) Performance

(7) Developmental potential

(8) Job endurance (career

orientation)

(9) Judgment, attention, memory factors

(10) Interrelationship of Judgment, Attention, and Memory

D. CURRENT LIMITATIONS:

Current limitations in selection technique and procedures result in misplacement and training failures and sometimes in weapons systems failures.

E. TRADE OFF:

1. Modification of task required of personnel: i.e., use black boxes.

2. Modification of current personnel limitations to make more highly qualified personnel readily available (short term trade off).

The time has now come to summarize this preliminary overview of the criterion problem in personnel selection. This can be done most effectively in terms of specific recommendations regarding research needs, programs, and efforts which are essential at the present time.

1. Current efforts to establish a **DATA BANK**, which would contain readily available criterion data from many sources within the Navy, should be expedited. The **DATA BANK** concept should become a reality. The compilation and utilization of criterion data from this source would be of great value in improving our selection techniques and efficiency, as well as in evaluating success. The **DATA BANK**, however, must be dynamic and constantly updated rather than static.

2. There is a need to standardize criterion information, selection data, and selection procedures. This is especially important as related to new weapons systems and new environments in which personnel are being assigned.

3. Additional research programs are needed to study the interactions **among** and **between** criteria.

4. Additional criteria data is needed to compare performance success of AFQT Category groups 1, 2, and 3 with those of Category 4, in view of recent requirements to utilize Category 4 personnel.

5. A major effort should be expended to develop adaptive classification criteria, to take

into account time, training, experience, and systems changes.

6. Temporal factors as related to the assessment of job proficiency and criteria of success should be given more consideration in selection programs.

7. There is a need to improve our criteria regarding the relationship between weapon system performance and crew performance capabilities.

8. We must identify and develop operational definitions of our criteria. Where possible, they should be situation specific as well as system specific.

9. We must improve our multiple criterion indices.

10. A major criterion development effort should be exerted in the use of (a) dynamic testing situations, (b) field testing situations, and (c) simulations. These, though task oriented, are needed for adequate selection and assignment of personnel to complex weapons systems. This should include both individual as well as team performance situations.

In conclusion, I would like to say that we have a marked need for "fine-grained" multiple criteria which include consideration of human skill requirements and capabilities, machine and situational requirements and capabilities, mission requirements, and those specific human attributes which are essential for successful performance within specific identifiable situations.

FREE INTERCHANGE FOLLOWING DR. CHAMBERS' OVERVIEW

CAPT CHRISTY: I'm sure we all realize the importance of this very comprehensive overview. At the outset, one broad class of problems in the human engineering area has to do with communications. How do you get engineers to communicate with medical-psychological people both interested in the same problems?

DR. WEYBREW: I have a comment concerning the changing nature of various systems (weapons systems in submarines for example), over the years. Dr. Haythorn mentioned crew effectiveness, group effectiveness and team performance and twice he mentioned

our inability to identify the crucial standards or criteria necessary for group assessment. I think this inability is becoming more and more evident, especially in submarines. An individual's judgment is affecting our weapons systems less and less. Rather, it is six or eight men, or interrelated individuals, making complex decisions in sequence or together that affect systems outcome. An important question therefore might be, "What methods or techniques do we have at present for gauging group effectiveness or team effectiveness?"

CAPT CHRISTY: One principle that we need to be aware of, particularly in selection for high risk duty, the astronaut program for example, is the notion mentioned in regard to the Pensacola program, the notion of sequential screening. After the man has survived officer training, Navy pilot training, a lot of operation training, and he gets to test pilot school and he survives that, he goes to Johnsville for evaluation. Up to a point, one can extrapolate from observations as to a man's tolerance to altitude, heat, tumbling, weightlessness, and acceleration. In talking to one of the astronauts, Frank Borman for example, he stated that he could go through simulation of a space mission for thirty days in a weightless condition wearing a pressure suit but at the same time he indicated he couldn't stand one "g" with the pressure suit on for ten days. This gets us into the problem of what can be simulated and, equally important, what can be done with performance data from simulated situations. I think we should try to simulate as many of these situations as we can, using man's maximum capabilities. Man, by and large, is fine in making decisions and judgments, but he is a lousy integrator of information of any magnitude. Here is where computers take over. In the past we've never had the centrifuge or other equipment or the test batteries or even the criteria when we needed it. Instead we took say a 100 men, or whatever number we needed out of what we had, and hoped they would do a good job. Many are reluctant to consider the problem of getting the people for SEALAB III or for space flight or for submarines or men for crews of other weap-

ons systems. We have to keep ourselves constantly oriented or we'll get lost in the laboratories and forget some of these interacting factors such as basic personality aptitudes, the stresses of the particular environment and the length of exposure to it.

DR. TOLHURST: I might make some comments about data banks to begin with because it seems to me that our weakness lies in the area of fleet performance or crew effectiveness information. It doesn't do a whole lot of good to talk in terms of selecting people that can get through the training program, if that's as far as we go. We have to get out to the fleet. I would make a proposal that we entertain the idea of a fleet performance laboratory charged with the responsibility to develop effective methodologies for getting such things as fleet criteria. We are hesitant to leave our home command. Often we don't have time to leave our home commands to go out in the fleet on an extensive data collecting routine. The outcome would be large amounts of fleet information stored in data banks. This activity could be on a carrier or a homeported ship making excursions to the operating fleet.

CAPT CHRISTY: Some of this is going on now. Several cruiser trips have been made, one aspect of their mission being to evaluate general weapons system effectiveness, to include personnel effectiveness, of course. We've run into the situation where scientists don't want to do this kind of research. They want to stay in their laboratories. Too, the operational people are too operation-oriented and too overwhelmed by the tasks they are responsible for, so that little help comes from them. We need more so-called human engineers. We don't have many of them. Most of them are working for NASA.

MR. MOLYNEAUX: It seems to me that there is something significant missing in the loop; it's as though we're talking about criteria as a word but we're not using the concept. I am of the opinion that the establishment of criteria is a command decision, purely, simply, and nothing else. The things that we're talking about are establishing characteristics of systems, establishing

parameters of systems—these aren't necessarily criteria. Criteria are those measures of acceptability which a system must meet in order to be utilized by the Navy. In the final analysis this must always be a command decision. In a smaller system the Command may be within the laboratory where development goes on. The command decision at that level is often up to the scientist, but for acceptability in the fleet it will, in the final analysis, be essentially a CNO decision. I repeat, criterion decisions themselves almost always involve command decisions.

CAPT CHRISTY: The trouble is often that it is difficult to get the data from the person who makes the decision. An example of another type of problem was during the Cuban crisis when it was discovered that some men were on 8 to 4, 4 to 12, and 12 to 4 schedules for two days at a time without taking into account disruption of biological rhythms. We were concerned about this. They had set it up this way because they were trying to be fathers and members of the community and yet hold a job in the command system; they didn't want to have watches of two, three, or four weeks and not see their kids and families. In short, it often happens that people are making decisions without adequate information, much of which possibly isn't out of the laboratory yet.

DR. TOLHURST: Let me develop an idea that was first developed by one of my cohorts, Dr. Glenn Bryan, also in ONR. He has suggested a useful concept, in my opinion. It is that the Navy should man and maintain, along with the rest of the ships in the fleet, a "people-ship." A people-ship would serve as a useful function in refresher training. It would also serve the function of having people available in the fleet, people who could do performance testing to include biomedical tests. At times it might serve as a quasi-hospital ship. I feel this is a fairly sensible idea.

CAPT CHRISTY: Except, where do you get the professional staff who are willing to leave their families or laboratories for long-duration missions? Too, fleet performance data will be very difficult to obtain on individual ships. We may not be able to test a man fully

on his job, say on a tin can, or on a people-ship for that matter.

LCDR NELSON: I presume we are talking about selecting observations in some way related to the criterion identification and development problem. We need to keep in mind that we must be selective in the way of observations, select what it is that we want to measure. In line with Mr. Molyneux's point (maybe we're dabbling with semantics, but I'm not so sure that we are), my recollection is that the first time that I heard of the concept of criterion was in connection with a rat maze study. We talked about whether or not the animal met the criterion. In this case the criterion was established as a standard or, if you wish, a goal. In this context, any system designed or conceived has at least one goal; it may have a hierarchy of goals. While no doubt the engineers have plenty of their own problems, it often happens that their logic is better than our own. We often delve too quickly into specific measurements, which in a sense may be operational definitions of how we reach these goals.

I'm reminded of when we started with the Antarctic work and what finally turned out to be the ultimate goal. The ultimate goal was to bring the men back alive. In fact, everybody did get back alive. Being interested in individual differences, we found little consolation from this information, but we felt successful. If we stop to think in terms of goal structures, often the same situation prevails when decisions are made with regard to trade offs. I haven't had much personal experience in this area; some of you here have. I understand that someone or some groups do make, probably increasingly so, decisions as to whether this system should be semi-automated, fully automated or a totally manned sub-system, or whatever. When such decisions are made, I would presume that they are made at least partially on the basis of information that either the machine by itself or the man interfaced with it or the man by himself, could do a better job in accomplishing the goal or mission of the system. This would seem to me to be the beginning of the rudiments of an outline to define the areas of performance that are relevant—in

what way is man superior to the machine? Why do we need man as part of the system? We are reminded recently that there's some criticism of the early Mercury pilots becoming political agents or public relations men. Well, I'm not so sure that perhaps man being in the system is not a function of some social goals and as a result it might be necessary that these people possess certain social skills. This kind of thinking seems necessary at this point.

CAPT CHRISTY: This is a good example. I was at Bethesda when we were first asked to screen candidates for Deep Freeze. Now we weren't given any criterion. We depended only on orientation. At first we were going to go to the Antarctic to get some idea of what we needed, adjustment to snow and ice and cold weather, possibly climbing mountains, and skiing. Needed possibly were some inner resources to read and study, and so on. It wasn't until we began to get feed back from the Antarctic study at San Diego that we had something to go on. One example of what can happen was the situation where an astrophysicist was screened out because of certain personality characteristics, more exactly because he was too much of an isolated, withdrawn fellow. What more could you want than someone who could get along well in isolation, someone who was going to sit day-in-and-day-out and take astrophysical measurements for a year at the South Pole? This goes back to the job description problem in one sense or the criterion problem in a broad sense. How do you dimension job satisfaction, anyway?

MR. MOLYNEAUX: One of the procedures which has become popular lately is Secretary McNamara's procedure of establishing criteria in terms of their so-called "cost-effectiveness." This is a concept we can use. Many believe that the cost effectiveness procedures have not been used in the psychological field as much as they should be. We go ahead and develop selection methods and techniques without looking at them to find out if the end product is worth the effort put into it. Would we be better off if we didn't do this selection in certain situations? Would we be better off if we selected some other way or with some

other goal in mind? I certainly feel we need to start looking for additional techniques to assess selection efficiency.

LT GREEN: Does anyone know who has worked out how to use cost effectiveness techniques in this context? This whole approach has had, and will continue to have, a dramatic influence on anything we do.

MR. MOLYNEAUX: I don't think there's too much of a methods problem if you know what you're attempting to achieve. This may sound like circular reasoning, but you start out with the process or, in some instances, the procedure you want to be cost effective. Then you look at the end result you want; then you try alternate procedures with "cost" data for each. All that is left is to compare outcomes and costs. Cost-effectiveness approaches have been worked out for many problems of this kind.

LT GREEN: It is hard to understand sometimes how you can actually utilize this approach in evaluating personnel management decisions; for example, problems of optimal utilization of skills in a context of severe limitations in available skills.

MR. MOLYNEAUX: Here is a simple example. Let's assume a selection system which yields correlations of, say, 0.20 with the criterion we're interested in. We're talking about 4% of the variance being accounted for by the selection technique. How much effort and resources does it take to have this productive capability? Say we're putting in 10% of the staff potential to achieve a 4% improvement in efficiency. Have we done anything positive for the Navy? And how much is it costing us in terms of money and other resources? This is a cost-type problem.

DR. TOLHURST: I think Dr. Chambers has given us some justification for the cost-effectiveness approach by his matrix (Page 39). But I think you need another dimension. For each of his dimensions of individual performance measures, you need another scale having to do with the range of sensor processes needed for effective performance. I realize all the complications involved in

obtaining this information, but it would appear to be a good place to start. A cost-effective approach could be applied also.

CAPT CHRISTY: This may be a trite example, but somebody in the Pentagon convinced someone else that a sizeable sum of money could be saved by cutting out fleet movies. But what would this do to the reenlistment rate of men on subs and destroyers for example? Movies haven't actually been "knocked off" in the fleet, but this is an example of the type of thinking which has become quite common these days.

MR. MOLYNEAUX: This is one of the problems any systems analyst runs into when he analyzes part of a system. He is taking a big chance of causing trouble in that part of the system that he didn't analyze.

DR. CHAMBERS: You recall that Dr. Weybrew noted earlier that the reenlistment rate of enlisted submariners within a given rating is higher than in any other sub-group within the Navy. It may be the sense of mission, leadership, personal motivation or something else that's involved. Presumably some of the submariner selection criteria should be more closely examined.

CAPT CHRISTY: You are dealing with a very special group. The motivation for them is more intense; their mission is seen as more important than most. I think the aviators are a very special group, too. An aviator actually doesn't cost us as much, even with pilot training, as a submarine officer does, for example. Submariners are expensive to train.

DR. CHAMBERS: I think one other point should be made. It is that the reason a man becomes involved in these complex weapon systems is because he is not only capable of making decisions but wants to. It is doubtful that most of the paper and pencil tests now in use get into this aspect. It seems clear that we need to get more into the area of management decision-making processes. What does it take in order to make decisions — absorption of information from displays, transmission of decision outcomes back to the weapon system, and so on. So communicative skills probably, and manipulative

skills are important. I don't think that there's any acceptable reason for us to wait until we develop good crew performance data in the fleet for us to start collecting data and deposit it in the data bank we've talked about. Five years from now we should have some useful fleet criteria as a result.

LCDR GALLAGHER: The question that comes up inevitably is, "Who is going to make the decision as to what data goes into the data bank?" For example, we have been talking about cost effectiveness which suggests a related question: "Are the cost effectiveness questions going to be answered by the investigator, by the operator of the equipment, by the test constructor and interpreter or by some high-level official?" Who along the way is going to make the criterion decision?

CAPT CHRISTY: There are two parts to cost effectiveness. The cost segment of it which generally is readily calculable and effectiveness part of it, often less easily estimatable. Command decisions become necessary all along the line in this type of analysis. We seem to agree that there is nothing like experience, that is military experience, for making certain types of decisions. But there is the ever present question of what constitutes effectiveness in a weapon system. It may make a big bang, but if it doesn't discourage the enemy, it's not effective. It may be quite accurate, but if it doesn't go off, it obviously isn't effective. To repeat, the problem is twofold: first, the Command side where decisions as to the most relevant goals or mission objectives are made; second, operational decisions, how can one operationally achieve these goals most effectively? Both aspects are important, it would seem.

DR. HESTER: I have what appears to me to be an interesting question. Suppose that we do specify goals. We'll make them quite limited. Let us say for example that our mission is to drop bombs within a certain area marked for destruction. This might be the mission of your aircraft as an example. Now you can presumably design a system which will do this. As Dr. Chambers indicated, however, you don't know how the people are

going to perform in it. But at least you can calculate in theory (this is where systems analysis is useful) what is the best possible performance one could expect from the system in question if there were no performance errors. One way of crew evaluation is indirect. Just see if the system meets your expectations. This gives you a device for talking about crew performing criteria without talking about how you got it. If the system doesn't operate as expected theoretically, then probably it can be improved in one way or another. But your cost effectiveness problem is something else when applied to this type of problem. Comparison or payoff data are needed before you can state which system obliterates the given area most effectively. These data then are essential before a new system can be constructed; most certainly before crew performance can be meaningfully evaluated.

MR. MOLYNEAUX: Where you have a man-machine trade off, inevitably you get into the cost/effectiveness problem. For instance, it may be considerably less expensive to put a crew aboard the plane and have a bombardier looking for the target, a pilot and a co-pilot and a navigator to take the plane over the particular spot. Now this may be the cheapest the job can be done. On the other hand, you may have some satellite navigation equipment and a vehicle either with a pilot or pilotless which can ride along a radio beam to the particular spot and dump the bombs on target more times out of a hundred than with any other system. Yet at the same time it may cost more to develop that system and to get acceptable reliability. This obviously leads to a cost/effectiveness decision.

DR. HESTER: To stay with our example, if you only have one area to obliterate with your bombs and the existing system will do the job in all probability the system will be more effective than anything you can design, but you have to project your needs.

MR. MOLYNEAUX: In actual operation, men have stress introduced; machines ordinarily don't. Machines operate within limits the same in practice as they do in actual operation. When malfunctions happen in

tests of machines, they are corrected and not likely to occur in operations. Others may appear, however. Obviously it is different when the "man-side" of the man/machine problem is considered.

It should be made clear that cost/effectiveness is a technique which is in and of itself something of a system for decision making. The implication is not that correlational and other techniques should be abandoned. It constitutes one of the techniques that validates your research. When your research provides information that ties into the decision process, then useful communication is established. So applied researchers should be trying to communicate with the Commands, to do the necessary research, the back-up work, the complete staff job, so that through the research data you communicate to the Commanding Officers, who in turn, are able to decide on the most expeditious way to get a task or mission accomplished. One of the things that we have not been doing is looking at the cost/effectiveness of our operations and proposals in order to give the Command the information needed for major decisions.

DR. TOLHURST: This is a rather complicated, often subtle, problem. Quite frequently the Command or those who are in the command decision position can't identify the most relevant questions to be answered. Often the questions are very general and hard to pin down.

CAPT CHRISTY: That's one of the reasons that we have a Submarine Medical Lab at New London and a Pensacola "Lab" and a Neuropsychiatric "Lab" at San Diego all located near training operational activities. In these situations knowledgeable scientists have excellent opportunities to become aware of the operational problems. Their scientific abilities can be put to use evaluating the various systems, or the man in the system, or the limitations of the machine and man. Then they should ask what they can do about the problems with which they come into contact. Of course, it is an endless process and quite often our data acquisition tasks are blocked, often on budgetary grounds, or in some instances, on the basis that the studies

may interfere with crowded operational schedules. Most of these barriers are surmountable, however.

DR. CHAMBERS: This would seem to be an appropriate note on which to terminate the discussion for SESSION II. We seem to have agreed that meaningful criterion data are most essential for any selection program though these data are often difficult to obtain, particularly under operational conditions. It would appear, however, that our discussions may possibly have served to refocus our thinking about this crucial problem area.

SESSION III

PERSONALITY AND APTITUDE MEASUREMENT IN SELECTION

SESSION OVERVIEW PERSONALITY AND APTITUDE MEASUREMENT IN SELECTION

Eric Gunderson, PhD
Naval Medical Neuropsychiatric
Research Unit
San Diego, California

LCDR Paul Nelson, in a recent paper*, suggested that selection should be viewed not as a single decision but as a whole series of decisions extending through a man's entire naval career. Dr. John Plag, in his presentation in the first workshop session, sketched some typical decision points in the selection process and indicated the kinds of information available to decision makers at different stages, beginning with the Navy recruiter and the AFES Station, through recruit training, assignment to a school or an occupational specialty, and finally re-enlistment or discharge. Sorting men into occupational specialties obviously is an extremely important step and for practical purposes an irrevocable one. Men are not only sorted in terms of aptitude scores, that is, Basic Test Battery scores, but at the same time, perhaps unintentionally, recruits are sorted on social background and personality characteristics as well. For example, in the Deep Freeze volunteer population approximately two-thirds of Navy carpenters had fathers whose occupations were in the skilled trades. In contrast, men in most other Navy enlisted occupational groups had relatively few fathers in those occupations. Hospital corpsmen, electronics technicians, and aerographers had high proportions of fathers in white-collar jobs than did other Navy enlisted rates. Actually, Navy enlisted occupations can be rank ordered on family socioeconomic status scores with aerographers and hospital corpsmen at one end and equipment operators at the other.

Associated with these social background differences are a number of personality differences. I'll not take the time to describe these relationships, as they have been discussed elsewhere*. I do want to make the point that as men are distributed throughout the naval organization by occupational specialties, they are at the same time being differentiated into more-or-less homogeneous socio-psychological subgroups by virtue of the selection standards and policies imposed.

At this point I would like to raise some questions concerning aptitude testing. Dr. Chambers implied yesterday that in view of changing and increasingly complex job requirements in the Navy, we may need a more highly differentiated aptitude test battery for selection and screening purposes. As I understand this point, Dr. Chambers was saying that for a particular aviation group in one training situation for example, testing of skills and aptitudes was necessary in order to place the graduates in the particular jobs that they could best perform. Perhaps in relatively late stages of the selection process, it would be feasible to develop a differentiated testing program for a particular area of work. Possibly the group can discuss later on whether, early in recruit training, we could obtain a better picture of what a man's future potential might be. After all, I think Dr. Guilford has identified 120 (or some very large number) ability factors. The sheer number, as well as the complexity of these factors, goes far beyond the bounds of any practical use. There is, I am sure, work going on to identify special aptitudes and abilities early in a man's naval career. But it needs to be remembered that selection or sorting of naval personnel goes on all the time, and, as a result, certain individuals may end up in highly specialized situations after passing many selection hurdles, such as those referred to by Dr. Chambers in his

*Unpublished manuscript prepared for the Third International Conference on Applied Military Psychology. London, England, 1966.

*Gunderson, E. K. Eric, and Mahan, Jack L. Cultural and psychological differences among occupational groups. *Journal of Psychology*, 1966, 62, 287-304.

overview of the criterion development problem. Other examples are found in the Polaris crews, naval aviation, and in Operation Deep Freeze. Incidentally, Dr. Chambers' method of representing the selection problem for a particular situation seemed to me very useful, that is, presenting the job requirements or criteria on one axis and the skills and attributes needed on the other axis. I noted that Dr. Weybrew, in one of his Memorandum Reports*, used a very similar technique for representing adjustment criteria and predictors for the submarine situation.

In terms of the selection principles that I've mentioned, at least for certain specialized naval applications, I think that we would agree that personality variables may have some value in predicting performance. The question is, "what personality variables are relevant in which assignments?" In terms of the grid model representing relationships between predictors and performance criteria, we would expect personality variables to be related to job requirements in a number of situations. Under what kinds of operating conditions would we expect personality variables to assume some importance in determining performance outcomes?

I would suggest two environmental conditions which might enhance the importance of personality variance: (1) extreme or unique situations which are quite different from those encountered in naval duty, for example, confinement situations; and (2) situations of prolonged duration. We are all familiar with the many studies in laboratories and elsewhere in which personality variables have failed to relate to any meaningful performance or adjustment criteria. I think that this is often true because of the short-term nature of the experience. If subjects have to perform certain activities for a few days or even a few weeks, personality differences, defensive responses, motivational changes, and so on, may not appear in such a short time; however, over periods of months, particularly in confinement situations, personal-

ity variables become more and more salient. The third category of situations in which I would expect personality factors or traits to assume considerable importance would be situations requiring close, continuous personal relationships within a group of heterogeneous individuals. For example, aboard ship we can identify certain work groups where people are constantly together, constantly communicating, making decisions, and responding to situations in a way that depends upon the group acting together. Other work groups and particular jobs do not function in this way. It seems reasonable to assume that personality factors become relevant as some complex function of the interdependence of the group members. The more personally involved individuals become, the greater the probability of encountering competitive, aggressive, or defensive reactions or other behavior that may prove disturbing to others within the group.

What kind of personality measures have proven relevant to Navy selection problems? I hope during the discussion period following my comments that many of you will report your experiences and perhaps recall personality factors which relate to performance criteria in your own specialized areas. Unfortunately, I am not familiar with many of the selection situations in which personality variables have been utilized; therefore, as a matter of convenience I will use as an example some of the data from the Antarctic situation that have proven significant. Three self-report personality measures have consistently differentiated men with respect to performance at Antarctic stations. As most of you know, assignment to an Antarctic station means being dropped off, for example, at the South Pole, in February or earlier. When that plane leaves, another plane will not return to pick you up again before 1 November. Meanwhile, there is no way to get into the station or to get out of the station if someone needs help; so it's a confined, closed group situation for six to nine months. The three personality characteristics measured by inventory scales are Achievement Needs, Activity Needs, and Affection Needs. Achievement Need is measured by the Edwards

*Weybrew, B. B. Some general comments concerning personnel selected for hazardous duty—after six years of research in submarine selection. SubMed-Res Lab Memo Report 62-5, 1962.

Personality Preference Schedule items, although we have modified the method of presenting these items. Activity Need is simply the number of hobbies and recreational activities that the person has indicated he likes very much. We assume that this reflects his need for such activities. Affection Need is measured by the Wanted Affection and Expressed Affection scales of the Schutz FIRO-B Inventory. These variables correlated with social and emotional criterion scores derived from supervisors and peer ratings. Dr. Radloff, you may recall, mentioned these findings in connection with the SEALAB studies. The explanation for these relationships appears to be that individuals with high achievement needs, particularly Navy men, are simply frustrated in this situation which frequently engenders passivity and low levels of accomplishment during the long winter months. Similarly, opportunities for recreational and social activities are severely limited in this environment and those habitually engaging in such activities are at a severe disadvantage. Individuals high in achievement or activity needs typically show increases on symptom scales reflecting depression, anxiety, irritability, and insomnia. Finally, with respect to the Affection Need variable, as the winter goes on men typically withdraw socially to some degree—perhaps becoming mildly irritable, critical of others, and less sympathetic and emotionally supportive towards others. A dependent person who requires much affection, attention, and so on, is at a great disadvantage in this setting. It is the last place in the world for such a person to be. We can see that particular characteristics of the situation interact with personality variables to produce decrements in motivation, performance, and personal or emotional adjustment generally.

Later on Bill Haythorn will discuss personality variables and group interaction, but I would like to make a few comments about group composition. We have measured the compatibility of Antarctic groups by means of a questionnaire administered at the end of the year. In these questionnaires we asked every member of the station to indicate on a number of rating items the extent

to which the group got along well together, was this the kind of group that the individual liked to be with, and so on. We used this measure as a criterion of group compatibility. We have related group composition variables to the rank ordering obtained on compatibility. One of the expected findings was that groups homogeneously high on a dominance measure, the Expressed Control Scale of Schutz' FIRO-B, were low in compatibility. Incidentally, this was a replication of an earlier study by Paul Nelson using a somewhat different measure of dominance. Results of a number of other studies, even short-term studies and such as those conducted by Bill Haythorn and Irv Altman at Bethesda, have provided similar results. If we place people with high dominance needs together in closed situations, we can predict that they will have considerable difficulty, probably early in the group experience.

Another personality variable, Autonomy, perhaps is a critical one for civilian members in Antarctic groups. Members of small closed groups who are unable to subordinate themselves to group norms and discipline probably will disrupt cooperative efforts of other group members. In any case, heterogeneity on the autonomy variable is negatively correlated with compatibility. There are other similar relationships which I do not have time to describe.

I would like to spend just a few minutes on some of the difficulties familiar to all of you in attempting to measure personality variables. It would seem to be a truism that measurement of personality variables is much more complex than the measurement of aptitudes and skills. I suggest that the principle reason for this (Dr. Radloff alluded to this in his SEALAB paper in the first workshop session) is the need for concealment. Personality variables can be estimated in several ways, but using self reports of personality dimensions involves particular difficulties. To what degree is the individual threatened in a testing situation? How do responses to particular test items vary with perceived threat, etc?

We have used several approaches to personality measurement. These are: self

ratings; ratings by close associates, that is peers; ratings by immediate supervisors; ratings by clinical experts, that is, psychiatrists or psychologists; and, finally, ratings of overt behavior by trained observers or objective records of behavior. Again Dr. Radloff gave us some clear-cut examples of objective measurements used in SEALAB II. In our studies we have found that self ratings had very low correlations with personality ratings by others. There tended to be good agreement between peers and supervisors, however. The fact that Dr. Radloff found in SEALAB II that peer ratings of behavior traits and performance were very highly correlated with objective records of performance tends to support our belief that peer ratings are a very reliable source of information about performance behavior. Personality ratings by psychologists and psychiatrists after approximately half-hour interviews have had low relationships with supervisor and peer ratings and very low correlations with self ratings. Yet self ratings, in certain instances, provide useful predictive validities. In short, we have a very complex set of relationships here in terms of the unique relevance of particular sources of information about personality to the prediction of behavioral criteria at a given stage in a man's naval career.

The type of objective data that Dr. Radloff described is extremely expensive and difficult to obtain. In most situations we must utilize much less difficult and less expensive procedures than prolonged observations by trained observers. Even peer or supervisor ratings may be impractical in many situations and the only data possible may be self descriptions. Thus, although more objective types of data are certainly desirable, we must also strive to increase the validity or the usefulness of self reports. This type of information can be obtained in almost any type of situation in which we are interested.

Let me suggest a possible approach to the problem of raising validity coefficients of self

ratings by taking into account the need for concealment variable. Couch* discussed this problem several years ago. Perhaps we could look upon the need for concealment as a kind of moderator variable. Persons high in their need for concealment might have a quite different set of regression weights and therefore warrant differential treatment of their responses with respect to predicting a criterion. Along similar lines, Dr. Hunt emphasized in his keynote remarks that the situational context is of very great importance, and we must look more carefully at how the individual perceives the situation at the time he is being tested or observed. For example, we were a little surprised that in the Antarctic situation our motivation measure at the time of screening turned out to be negatively correlated with performance in the Antarctic. The motivation score was the individual's expression of how much he wanted to participate in this assignment. One assumes that in most job assignments the man must be positively motivated and in a difficult situation perhaps even highly motivated. In this case high motivation scores were not predictive of success. I think that this may be related to honesty or realism in response to the questionnaire. We consistently find that men who express a little scepticism, or perhaps realism, about their own jobs, their own personalities, and the rewards they can expect in an Antarctic expedition tend, in general, to show more adequate adjustment to the environment.

I'm afraid that I'm going to have to leave the question of personality measurement right there and ask for group discussion about handling the problems of the defensiveness of the subject and his perceptions of threat in test situations and the utilization of self report measures more effectively. Incidentally, I don't think that the Lie Scale and the K Scale in the MMPI have solved this problem.

The issues of using self report inventories also is of special interest in the area of symptom description. At the Neuropsychiatric Research Unit we are working on the problem of evaluating and predicting psychiatric illness and other illnesses. Is a man's

*Couch, A. S. The psychological determinants of interpersonal behavior. Proceedings of the XIV International Congress of Applied Psychology, Copenhagen, 1961.

own report of his health status a good indicator of whether he is sick or not? Does a man's report of his health status agree with a doctor's appraisal of his health status? This question is a critical one inasmuch as self reported symptoms are used extensively in evaluating health. Regardless of state of health, there is a positive correlation between self report and doctor's report, of course. There is generally less defensiveness or concealment with respect to the physical symptomatology than psychiatric symptomatology, but we have to deal with the same issues mentioned earlier in the context of evaluating health status.

I want to take a few minutes to refer again to the central issue raised during the Session II discussions, namely the problems associated with specifying organizational effectiveness criteria. Ultimately, we must relate selection procedures, personality measures, training, and all other factors to the final product, namely, the operational effectiveness of combat-ready units. How can the operational effectiveness of units or organizations be measured? A psychologist who had training duty with us recently in San Diego described an interesting procedure for evaluating organizational effectiveness. He had participated as a research observer in the study designed essentially to assess the combat readiness of ships before they were deployed to the Western Pacific. In brief, every naval vessel before leaving San Diego harbor, goes through a series of simulated combat actions involving responses to submarine and air attacks, damage to the ship, and all emergencies that might arise. Simulated emergency situations are presented in a pre-programmed and reasonably standardized manner. The performance of the entire ship is scored objectively. If the ship receives a high enough score to pass the test, it is deployed. If it does not pass the test, it is not deployed.

The responses of the ship, such as evasive maneuvers, firing missiles, or whatever, are timed, recorded, and evaluated against a standard. Breakdowns in communication, delays in transmitting or executing commands, failures or deficiencies in equipment,

individual incompetence, and poor teamwork often become glaringly apparent. It is suggested that useful kinds of organizational criteria may be found in this type of evaluation of operational effectiveness. The end results of selection and placement in terms of the readiness of every man to perform his duties, to communicate and cooperate effectively, and to perform as a team member are all represented in this kind of organizational evaluation. One could look at many aspects of the organization's composition and role structure that contribute to the effectiveness of the total operational unit.

In San Diego we are especially interested in certain aspects of organization and the incidence of illness. At the moment we are struggling with the problem of identifying ecological, organizational, and situational factors aboard ship that are related to illness incidence. We have identified certain significant factors, for example, the man's status or pay grade, his particular job and division on the ship, his level of job satisfaction, and a number of demographic and social background factors. These variables relate to the incidence of sick calls and sick bay visits occurring during a given period of time. Furthermore, with respect to the ship as a whole, the sick call rate is related to the operational activities of the ship. When the ship is going from San Diego to the Western Pacific, the sick call rate is 3.5 per day. En route to or from port it is about 4 per day. En route to participate in combat operations, the sick call rate almost doubles, going up to about 8 per day. During combat, the sick call rate drops to about 5 per day. Returning to the U. S., that is, coming back home, the sick call rate drops almost to zero. There are some very difficult problems ahead in attempting to relate the effectiveness of total organizations to all components of the ship and to individual performance. As we continue to collect data aboard ship, we will try to develop a framework to accomplish this. Incidentally, it is a very great advantage in this effort to have a psychiatrist on our staff, Dr. Richard Rahe, who spends much of his time in the field aboard ship collecting data.

I think I should stop here and begin to get your contributions to the question of situations in which personality factors are important. How can personality traits or dimensions be measured? How does such measurement relate to various criteria of individual and group performance? Also, I would like to keep the question open concerning aptitude measurement and further developments in this area. I am not competent to discuss aptitude measurement and must rely upon your comments in the discussion period.

FREE INTERCHANGE FOLLOWING DR. GUNDERSON'S OVERVIEW

DR. TOLHURST: I think the point that Dr. Gunderson brought out is the point I intended to make a little later today. It is that in personality testing and, of course, biomedical testing, we have to ask the question, "Testing for what? For what purpose." It's important to have all the information possible on a man, to collect a variety of data.

DR. RIMLAND: I was interested in the amount of emphasis that Dr. Gunderson gave to what he referred to as self-concealment. Apparently he believes that this factor is one reason, perhaps the primary reason, for the ineffectiveness of the self report. I frankly am rather skeptical about concealment being the major reason for the insufficiency of these kinds of tests. Perhaps you mean concealment in some different way than I interpreted it, but from various bits of information that I have, for example, from studies involving the comparing of scores from people who took the tests anonymously as compared to those obtained from people who sign their names, bits and pieces of information like this suggest to me that the invalidity of self-report measures, to which I will in general agree to, is not traceable in a large part to concealment. I think there are other factors involved. I don't propose to be able to explain what they are, but I really am skeptical about self-concealment being the reason for this dilemma.

DR. GUNDERSON: I would like Dr. Rimland to react to this idea. Do you think this differential (and often quite low) validity possibly results from differences in subject matter

with which tests are dealing? Might a person be quite honest in rating himself, yet be able to rate himself more accurately in certain content areas than in others? As a result, shouldn't we expect higher correlations with our criteria for certain kinds of personality data? I don't have a specific example in mind, but I think this is an important issue.

DR. RIMLAND: Well, I've been rather impressed with some of the success that we've been having lately in the Navy Personnel Research Activity in San Diego through the use of the Strong Vocational Interest Battery. I have also been impressed with the published validity for the "Strong." For whatever reason, items which have the kind of item content that the Strong Test has appear to have useful validity for a number of purposes. This is a circular statement, but nevertheless, when we're talking about content of personality tests, then items like the Strong should be included in some of your selection batteries. I'm not sure why; it is possibly the manner in which the Strong was constructed and keyed, rather than the item content per se, that makes it useful. I don't know.

I will say that some of the other tests that we've played around with, some of which we've constructed and others which we've borrowed elsewhere, many of which have been based upon different kinds of test construction ideas such as forced-choice and so on, don't work nearly as well.

DR. HUNT: I'd like to contribute briefly to the concealment problem as it relates to the self report. What is going to happen to the concealment variable as our culture, as a matter of fact, accepts subjective personal test data of this kind? I think invasion of privacy is a consideration, but I feel people will eventually be willing to accept this. I wondered if you'd thought of manipulating the test situation and perhaps starting with your test instructions something like this: "We, of course, have a great deal of material about you. You know we have cleared you through security. We know a lot about you. This test is a self-insight test to see how well you can recognize certain traits within yourself." I've no idea what would happen. Anyway, I've been thinking of this for the past

couple of years. It would be an interesting idea, I think, to try.

DR. GUNDERSON: I think that touches on a critical aspect as I see it, how to reduce the perceived threat. You could point out that you might as well conform because there's really no need to conceal information, or you might mention that nothing is to be gained in a selection situation by concealment. Dr. Haythorn and Dr. Altman of Bethesda have been working with a self-disclosure scale in an effort to ascertain to what degree the individual will reveal information, intimate information, to other people. Now working at the other end of the problem, rather than the need for concealment we can look at willingness to disclose about self. Certain of the data they've collected I think is going to be very helpful in suggesting under what conditions people will or will not reveal certain personal information. Perhaps Bill Haythorn will comment on this later on.

DR. NELSON: I'd like to reinforce what has been said earlier in the Workshop in part by Dr. Radloff and in part by Dr. Hunt. Addressing myself to the problem of the need for greater focus on situations again, I'd like to see more original work done with biographic and demographic information. Retrospectively we have the great figures of history documented. We often refer to these individuals as rather classic standards of validity for some of our profiles of personality or whatever we have. It's interesting that none of these great figures has ever been described in terms of any of the personality tests which are available to us today, but instead are described anecdotally through the characterization of how these people behaved in life. We talk to one another to get impressions of one another; we go to peers with a great deal of confidence in peers' ability to evaluate one another and form judgments in a variety of situations. What is it that we're tapping in these sources of information? Admittedly we've had limited success with what we've done so far with demographic and similar

data. It may be because we've shown a lack of originality. In this regard I think the sociologists are ahead of the psychologists. Interpretation of the context in which past events of a person's life occurred are important. We know within the immediate setting, for example, that behavior in high school, extra-curricular and curricular, is related to the adjustment of the enlisted personnel in the Navy. There have been several books dealing with the subject, one of which I think is particularly interesting and written by James Coleman, entitled *Adolescent Society*,* a comprehensive study of a sample of high schools in the Chicago area. Here we find again, and it shouldn't surprise us, that nobody has really done a great deal of work in this area. Within different high schools the cultures are different, the norms are different. The meaning of being a captain on the football team is different in one high school from being in the National Honor Society, and so on. Again, I think we should expend more effort upon an assessment of past history. We've talked about demographic, generally we call it biographical data, to include those experiences or behavior manifested prior to the man's coming into the military. The fact is that the man's life continues in the military. Everything that happens to him, and that he does, constitutes for behavioral scientists, I think, demographic data on a continuous basis. Sometimes I think we get into the problem of restricting ourselves by the arbitrary constraints which we impose on behavior. We say, for example, we're in a procurement or initial selection process at the moment and within the next three days we will begin training. Now a man is in a training phase and later he may be in advanced training, then in a shakedown period and finally he is on his first assignment. What is happening to this man all the way through these experiences? He's behaving and he's responding to a variety of situations and as a result I think perhaps we ought to give a little more attention to the assessment of behavioral potential through what the man is actually doing and what he has been doing. Incidentally, I agree with Dr. Rimland's comments about the value of interest inventories

*Coleman, James S., *Adolescent Society*, 1961, Free Press.

and the like. There also may be other personality dimensions equally useful for our various programs.

Finally, one other comment I would like to make is that it seems to me that relatively few people in the selection field have given much attention to the structure of personality as such in developing their programs. We generally define personality, I guess, as being related in some way to what might be called behavioral orientation. This concept presumably bears some relationship to his attitudes toward his work or other activity and to his family and to other people and to himself for that matter. Yesterday, Dr Hunt mentioned Dr. Eysenck's* study showing one essential factor running through a sort of extrovert and introvert concept, namely, emotional stability. We know people demonstrate these orientations or so-called life styles differently. One of the concepts or constructs which I have been interested in, I know Dr. Haythorn and Dr. Altman have been too, is the concept of rigidity. It's a sort of integrative concept that ties things together. I don't see a real breakthrough here, but perhaps there are ways we can approach this type of concept through analysis of demographic and biographic data.

DR. O'CONNELL: With the midshipmen group at the Naval Academy, there seems to be a relationship between the socio-economic background and the way they can relate to the group, depending on whether their family came from the trades, or managerial or professional or military backgrounds. Too, there are wide differences in the sick call frequency depending on the structure of personality, participation in athletics, knowledge of sports, and so on. We know something about the relationships of certain behavior seen at the Academy to biographical and other personality data. Also, I mentioned in the first Workshop Session, knowledge of prior skills at the high school level seems to be related to success. Finally, some midshipmen conceal their lack of motivation by illness over a long

period of time and then when they're finally told that they can't get out this way, they say, "I really don't want to be here," and then they resign.

DR. INMAN: One thing seems certain that we are touching on a very controversial area when we talk about the concealment problem in personality evaluation, broadly conceived to include personality tests, projective and objective as well as interview approaches. One point should be made, namely, one can develop such a suspicious attitude that he may discard perfectly honest information as fabrication. We need to analyze the total situation in which the data are collected, for example, what are the "stakes" involved if something socially undesirable is admitted.

DR. WHERRY: I must confess a certain disenchantment with personality tests. It has been my observation that the universities are able to manufacture tests faster than we users can try them out. Having used hundreds of them at Pensacola and having seen other people use hundreds of them in one situation or another and having been unable to add to the multiple "R" by using these test scores, I frankly have lost confidence.

DR. GUNDERSON: Do you have any comments on simulation testing with respect to stress as an approach to personality assessment?

DR. WHERRY: We're attempting now to assess personality factors in terms of how the man reacts to real life situations instead of how he reacts to paper and pencil tests. The approach looks promising. For example, we are currently conducting a study in which the man is placed in a situation in which he is threatened with a weak electric shock. People react differently to these situations. At present, we are taking measures of performance degradation in this situation. The subjects generally seem unable to process information as accurately as they approach closer and closer to possible electrical shock. This effect is seen in particular three minutes before the electrical shock or immediately after the electrical shock or immediately after the time the likelihood of receiving electrical shock was maximum.

*EYSENCK, H. J., Classification and the problem of diagnosis. In H. J. Eysenck (Ed) *Handbook of abnormal psychology*, New York, Basic Books, 1961, pp 1-31.

I don't know if it is possible to develop situational tests to get at most of the concepts used by personality theorists these days. I find myself going from one of these concepts to another trying to anchor on something that will be predictive of later life adjustment situations of interest. I frankly find myself at a loss to know what direction to take.

Finally on a methodological note, I do have a few comments about the forced choice response approach. I think forced choice does offer some possibilities, but it is crucial that the item choices are equated in terms of relevance for the respondent at the time the data are collected. Also in Navy selection, the items need to be equated as to how important the Navy thinks this or that item is, as well as how important does the man think the item is.

MR. PARKER: One of the major problems in personality measurement probably recognized by everyone in the field is how do we validate these observations and measures collected by means of tests of various kinds. How do we find which personality characteristics or types are relevant for adjustment to submarines, space flight and other unusual environments? Of course, we come right back to the criterion problem. Until we can come up with suitable criteria for the situation we are trying to predict, we will be lacking direction in our research.

LT GREEN: I have one question and one comment. The question is: "what are we calling personality?" There are undoubtedly many definitions. One of the things I learned as a graduate student about personality is that the term refers to dynamic processes and, as a result, while collecting test data from someone in a given situation, you are at the same time getting information pertaining to a person's characteristic style of life, his "fight-for-life" personality. I don't think that all of the negative findings already alluded to in this Workshop session should be taken as a complete condemnation of the personality field in general or the personality concept in particular. What we may need are some new approaches, possibly aimed at purging the field of some of the things done in the

academic area over and over again. We need new bold approaches in the field.

DR. CHAMBERS: If I understand Dr. Hunt, it seems to me that he is suggesting an interesting approach. It involves focusing upon the nature of the adjustment situations as well as the personality in order to obtain a sample of the behavior or activity in question. We have tried this approach with some success. One problem arising in this context is that you may focus on the wrong situation in which to observe personality in action, so to speak, and as a result, criterion validity suffers.

CAPT CHRISTY: I agree with several previous comments regarding the usefulness of data pertaining to a man's activities, his performance, if you like, day in and day out. Of course, we must realize the time and effort involved in a detailed examination of each man's "jacket." Obviously this approach is prohibitive for large selection tasks, but for smaller special assignment screening (Sea Lab or Space Flights) there would appear to be no acceptable excuse for not doing something like this.

LCDR GALLAGHER: Years ago, I remember studying a research proposal involving large photographs of the face. By looking at the lines in the face, you could predict what the subject would look like fifty years from now. I had no idea what this had to do with the individual's effectiveness, but it did predict something. So the relevance of our predictions needs always to be considered.

Another thing I'd like to comment on has to do with this so-called situation approach we've heard so much about in this session. Predictive techniques which work in some specific situations, apparently don't seem to work too well when the situation is changed. Supervisor ratings might be a case in point. Those were tried recently in SUBPAC. The variance was narrow; everyone was a good guy. Another example from aviation was a pilot who, through a performance error, killed himself. I thought for sure this would throw off his crew ratings. Instead, they put a "halo" around him and he became one of the "best" fellows in the group. So, supervisor

ratings in this situation at least, were of little use to us.

DR. PLAG: Generally speaking, our experience at the Neuropsychiatric Unit in San Diego argues that personality test information per se really doesn't add very much to our predictions. The important thing is not whether this kind of data is predictive, but whether it adds anything to what you already know about a given man. I do feel that we should, in our thinking, separate measurement of personality as such from measurement of illness—personality illness. Are we measuring personality or personality illness? I'm not quite sure in some selection programs in the Navy. Valid assessment of the proneness for personality illness might be a useful focus and could add weight to our prediction formulas. Also, are we measuring personality, or something else, when we consider how the man performed prior to the time he came in the service, his level of schooling, his achievement quotient from school and this type of information? I showed in my comments in the first session of this Workshop that data of this kind are predictively helpful.

DR. HAYTHORN: It seems to me that attempting to predict performance from personality measures in a multiple "R" sense isn't going to get us very far. But what may be a fruitful approach is to look at the interactions between personality and aptitude variables. In support of this, Dr. David Kipnis, when he was at the Bureau of Naval Personnel and following that when he came to NAMRI and Project ARGUS, has done some sound work.* He developed two tests, one called a test of Persistence and the other a test of Insolence. Then he divided his population of recruits into four intelligence levels or groupings on the basis of the GCT. We then followed these men up eighteen months later in their Navy career. With supervisor ratings as criteria, he predicted the relationship between these personality measures and performance within each of these four intelligence levels separately. What he found was,

that among the "dumb" kids the correlation between Insolence and performance was zero, but as intelligence increased, the correlation became increasingly negative. Among the top quarters of the intelligence distributions, negative correlations in the neighborhood of -0.65 were found between Insolence and Navy performance. For the Persistence scores, similar findings, but in the opposite direction, were reported. Among the least intelligent, he got a fairly strong positive correlation, in the neighborhood of 0.60 , between Persistence and performance eighteen months later. This relationship became increasingly lower in the higher groups, until the brightest group, the relationship was zero. You might say as long as he has the "smarts," he doesn't have to have any persistence. I think the point of relevance here is that the relationship between personality and performance is not accurately described by linear correlation techniques. One has to take into account the interaction between personality and ability, an approach not seen too often in the literature.

Dr. Kipnis also did something else with this Insolence scale. He took the view that not only do you need to make a decision to select or reject a man, but you also want to know, if you accept him, what's the best way to use him. Here again, he found that his high Insolence subjects responded well to very strong punitive leadership and very poorly to the democratic permissive leadership popular these days. The opposite pattern was found for the low Insolent group. I would like to see approaches like this one followed up.

Finally, I have a lot of faith in crew composition as a technique that may eventually help us to select more and more effective Navy crews. Complicated problems are involved however. To illustrate one complication based on probability theory, the number of combinations of "n" men with "k" men per team is $n!/(k!(n-k)!)$. So, if you're assembling ten three-man teams, there are upward of a trillion different ways you can do it. More important perhaps, is the fact that we don't have the proper mathematical models for relating individual scores to team scores. We

*Kipnis, D. The relationship between persistence, insolence and performance as a function of general ability. *Educ. Psychol. Meas.* 1965 (b), 25, 95-110.

need a lot more effort directed to modeling problems of this kind, in addition to getting more information about the content and effects of crew composition.

DR. HESTER: There is one point on the criteria problem that interests me. Sometimes I think our notion of scored tests measuring a distributive trait over a population may be a major source of difficulty. What I mean is that a personality test is supposedly standardized over a population. Yet, for example, on a symptom-type test, you might have a very meaningful device for measurement of a population with respect to something like ego breakdown, whereas, the same variables mean practically nothing when you are dealing with the question of high level performance. If this is correct, then in the SEALAB situation that Dr. Radloff was talking about in the first session of this Workshop where the concern is with who is better and best, a measure may have limited value; whereas, if you're trying to distinguish between fair and poor, it may be quite useful.

DR. CHAMBERS: I'd like to add to what has already been said about the use of operational situations in personnel assessment and selection. In the early fifties, we used to go out to the personnel "pipe line" and give oral tests to mechanics. The problem was to select Air Force mechanics for certain bombers, the B-29 and possibly some old B-52's. The approach worked out pretty well. I was impressed in those days with the absence of correlation between the oral evaluation of the mechanic, i.e., on-the-spot question-and-answer examination of the man in close proximity to the engines he would be expected to repair, as compared with the written mechanical test. It turned out that mechanics as a group do not take tests very well. In fact, many of them were quite bored by taking tests. But to stand out there with your tape recorder and get the man's responses on the spot proved to be a useful selection technique, though obviously quite laborious and time consuming.

Later on, in the area of simulation, I arrived at some similar impressions. For hours we watched people sitting in trainers for example, going through procedures and tests

of various kinds. But they do a lot of things besides performing the tests we ask them to do. For example, they might demonstrate some idiosyncratic approach to problems that might be scorable, possibly weighted some way and used in the selection programs. The Navy has a number of trainers designed for many purposes and which conceivably could be used for this purpose.

For example, several years ago, I was being tested against a chimpanzee. You may know there have been hundreds of chimps tested at Holloman Air Force Base. In this case, the chimp was to make a response with his right hand every ten seconds and perform another operation at the same time. In one study, the specific task was to push a button with his right hand when a particular light came on, in order to avoid an electrical shock of considerable intensity applied to his feet. We were interested in whether a man or a chimp would perform better in this situation. It was interesting to me that the chimp was very proficient in operating the controls. However, one of the technicians happened by with a pair of pliers, whereupon the chimp started screaming and yelling and disrupted everything. When they gave the chimp a pair of pliers to hold in his hand, the test continued uninterrupted again. This was an example of an individual reaction, which could be measured and included along with the performance data. This interaction of unique personality traits with performance may be quite valuable in certain selection situations. Incidentally, the chimp actually beat me, except during the pliers incident, when I beat the chimp.

DR. GUNDERSON: You weren't disturbed by the pliers, then?

DR. CHAMBERS: No, I didn't care about the pliers and that's exactly the point I'm making. Some do and some don't show severe emotional reactions under stress. Sometimes something that's quite incidental to the real task upsets certain predisposed persons. However, the consequences of the incident may be quite relevant for mission success.

DR. MOONAN: I'd like to go back to an earlier statement made in this Workshop

Session, namely, the comment regarding the relationship of motivation to performance. Apparently, these relationships, while usually positive, are not remarkably strong. Too, under certain conditions, the relationship is negative. In light of some work we've been involved in in San Diego, I would expect that motivation, at least some assessment of it has been, and is being, used to facilitate the assignment of recruits to recruit training centers. Quite possibly, these interactions of motivation, aptitudes and other variables should be examined more closely.

I have not "lost the faith" in personality tests for the simple reason that I never had it to lose. I feel that it might be inappropriate to use such measures in any predictive system. I believe a man carries along not only his innate abilities, his acquired skills and his physiognomy, but also his personality into the adjustment situations we are interested in. I'm not advocating abandonment of the use of this kind of information. It may be that the approach is relatively ineffective because the art or science of measuring those variables is not advanced enough. However, my work as a statistician leads me to concur with Dr. Wherry that the contribution of personality test scores in predictive systems has not been very great.

Finally, I'm sure that all of the participants here know that the interaction effects which have been mentioned by several people in this session of the workshop can be separated statistically by multiple regression techniques. My experience has been that when these interaction effects are delineated, the results, characteristically, are not spectacular in most circumstances.

DR. GUNDERSON: I can report that in another selection setting, namely, selection for underwater demolition training, expressed motivation is very important in predicting whether a man will stay in the program. The approach appears to have useful face validity also.

DR. RADLOFF: In the SEALAB study the relationship was zero, using the same measure. So here are examples of situations in which the same measurement technique

bears opposite predictive relationships with the available criteria.

MISS AMBLER: Our pay-off at Pensacola has come from an empirically-validated questionnaire approach based largely on non-cognitive material that appears to augment aptitude measures. The fact that we have had some degree of success with this approach forces me to agree with Dr. Nelson and others who have argued for more emphasis upon demographic and biographical data of various kinds as part of an on-going personality assessment program.

A biographical inventory approach shows promise in our program particularly at this time, possibly because of the nature of input population. I use the term "opportunistic" to describe the men coming into aviation. The draft pressure is on and a sizeable number of capable men are looking for an easy commission. I hope to have some data soon to show that increasing the weight of the biographical portion of our selection battery will have a stronger predictive effect than increasing the entire cut-off score for the total battery which includes the aptitude. There is no gain in predictive power from raising the aptitude cut-off level, yet, there is some gain realized by raising the non-cognitive level.

I'd like to mention some experiences we have had with the so-called concurrent and predictive validity of our battery from time to time. Many of our "trial" personality scales will give useful concurrent validity after the decision to drop has been made or after training anxiety has developed. But we are not able to turn the clock back and pick this validity up in the predictive situation. I don't know whether this results from self-concealment, unreliability of measurement, or change in the individual. I think that perhaps all three are operating.

Finally, our experiences with interest test data obtained from the Pensacola population is that, when variance due to abilities, is partialled out, the predictability of the interest measure goes out too. Of course, our population is unique in many respects as compared to some of the other populations

sampled in the selection programs of the other activities represented at this workshop.

DR. WEYBREW: Dr. Wherry has indicated his disenchantment with personality tests generally. Some time ago, I became similarly disenchanted with linear regression techniques generally as applied to most personality test data. I frankly think the linear model is much too simple. Our experience in submariner selection, at least when wide variances in our criteria are available, is that the predictors do not bear linear relationships with these criteria. Our findings argue for configural, non-linear approaches. As approximations, simple cross-breaking techniques which use first one then another variable(s) as moderators will show up your gross predictive interrelationships to be followed by more detailed itemetrics and predictive equation formulation later on. If I understood Dr. Kipnis' approach, described so well by Dr. Haythorn a few moments ago, he argues for the same thing, namely, for the appropriateness of what has been called by some, the moderator-variable approach.

It was my idea to put aptitudes in with the personality session of the Workshop. I plead guilty. I did this advisedly. It seems a truism that aptitudes are specialized abilities and they can be correlated with skills attainment, most often defined by training criteria. This predictive capability should not surprise any of us, since aptitude test construction usually involves achievement indices (grades for example) as criteria. But this correlation between aptitude test scores and achievement criteria is characteristically found when the ability is being used under ideal conditions. What I am looking for are strong predictors of functional capacity under less than ideal conditions. I want to identify and grade individual differences in applied abilities to operate an escape hatch, to fix a transducer and so on, performance upon which the survival of the submarine or

spacecraft may depend. Here is aptitude that counts. Our data support the repeated finding in the literature, "that the more remote your criteria are from the training situation, the less the predictive power of aptitude test scores." Here in my opinion is where the aptitude-personality interactions that Dr. Haythorn and Dr. Kipnis talk about come in. Dr. Wherry's behavioral indices of individual differences in reaction to laboratory-contrived (shock-threat) stress may be "tapping" personality dimensions that could conceivably contribute to this interaction.

Along somewhat the same lines, some of you may know of our report* of the predictability of certain peripheral indices of autonomic nervous system (ANS) function with respect to submariner adjustment criteria. I hasten to add, however, that these findings have never been cross-validated or revalidated for that matter with trustworthy criteria. Yet, I for one, think these ANS indices of emotionality operationally define an important aspect of personality.

Finally, is there clear-cut negative evidence in the literature that some of the so-called perceptual measures do not interrelate with other personality dimensions in a meaningful way? I have in mind such measures popularized by the Gestaltists such as figure-ground reversal rate, closure, hidden-figure tests, illusion latencies, and figural after-effects to name a few. Intolerance of ambiguity, level of adaption and the old Lewinian concept of level of aspiration also occur to me as examples. Have we exhausted these ideas?

DR. RIMLAND: I would like to mention a finding from our BUPERS San Diego Laboratory. Dr. Ed Rundquist on our staff there asked me to read a paper he had prepared dealing with the validity of negatively worded test items. I was very impressed with the powerful evidence that he presented favoring the idea that negatively worded personality items are much more valid than the usual kind. Some of you might regain your interest in personality tests (which have really taken a beating in this Workshop Session) if you take a look at these findings.**

*Weybrew, B. B. Selection of men for hazardous duty from indices of individual differences in Autonomic Nervous System reactivity. USN Sub-Med Cen Report 65-1, 1965.

**Rundquist, E. A. Item and response characteristics and personality measures: A reaction to L. G. Rorer's "The great response style myth." *Psych. Bull.* 1966, 66, 166-177.

DR. MOLISH: After hearing all that has been said about the personality field, one might suggest that the most qualified person to summarize this session would be a minister trying to restore the faith.

I might make a statement that personality tests are not dead, nor is Freud dead, nor is the criterion dead, though the shadows may be lengthening for them.

With apologies to Dr. Hunt and many of my colleagues who have heard this from me before, namely that personnel selection for military duty dates back to the old Testament to Gideon, whom you may recall, selected his foot soldiers from their response styles "of drinking water." One cannot help wondering what criteria were involved. But simplicity characterized his technique. Is it possible that our methodology is too complex?

One prime point raised during these discussions would seem to involve the notion of temporal resampling of selection data. This emphasis makes the assumption that the selection process for the enlisted submariner for example, extends from boot camp, through Basic Submarine School, submariner qualification and finally, to the re-enlistment decision point at the termination of enlistment. The key concept here is a periodic re-evaluation of the man's general adjustment status. One important personality dimension or possibly, "family of dimensions" is motivation. Not only the assessment of initial motivation, but also the problem of motivational maintenance seems crucial for most selection programs. For example, in the context of the longitudinal

approaches I have just mentioned, how do we detect subtle (or often not so subtle for that matter), declines in motivation at selected career levels? Another interrelated point has to do with what might be called milieu assessment to include not only within-service assessment but also assessment of the quality of his extra-Navy (marital, civilian, community) adjustment as well. This emphasis makes the rather obvious assumption that a military man is one personality when on duty on-board the ship and quite another when he assumes the role of husband, father, and often, a military resident of an essentially civilian community. The total milieu concept demands also that periodic evaluation of the adjustment status of a person must take into account life situation changes, such as marriage, divorce, deaths in family, even perhaps the fact that he has a new commanding officer. At least in our submariner population, we have seen numerous instances wherein these complex factors have contributed materially to a man's psychiatric status at a given time.

I agree also with Dr. Rimland's favorable comments about interest tests generally, in particular the Strong Vocational Interest Battery. It may be worth mentioning that Dr. Weybrew's Staff at the Submarine Medical Center currently is engaged in an ongoing study designed to identify those factors affecting the career choice of Submarine Medical Officers. Certain of the Strong subtests are included in this test battery.

DR. GUNDERSON: Our time has run out. Thank you for your most insightful comments.

SESSION IV

BIOMEDICAL INDICES IN SELECTION

SESSION OVERVIEW BIOMEDICAL INDICES IN SELECTION

Gilbert C. Tolhurst, PhD
Office of Naval Research

The rationale for a discussion of biomedical selection and the biomedical aspect of selection in a medical institute or medical research laboratory for someone who calls himself a psycho-acousticologist is a very peculiar type of assignment, but I'll do the best I can. Somewhat in the way of apology, I'm not going to be able to offer very many positive suggestions; I'm merely going to ask a lot of questions. I shall present very little data and, unfortunately, even that which I mention will not be my own unlike the speakers who have proceeded me.

In trying to determine a logical approach to the topic of Biomedical Selection Procedures, three major areas become apparent: the current selection tests; tests which might be added and are used for certain peculiar or unusual environments; and then a listing of a group of other tests or measures which may be useful in terms of some published laboratory data or which may have some assumed relevance as a result of the circumstances under which the data are collected.

Before talking about the three aspects or broad general areas, may I applaud the organizational matrix developed by Dr. Randall Chambers in his overview of the criterion problem in the second Workshop Session. You may recall, there were categories of operational tasks along the horizontal axis. The "rows" axis listed factors which were involved in the task, sensory, perceptive, cognitive, and others. Similarly, I would like to be able to provide a catalogue of the various biomedical indices; however, time does not allow for more than a start on such a task.

For most military situations it is comforting to know that a man is healthy. It is self-evident that if a man has a certain

degree of health, his chances for survival are better than if he does not. However, some of the criterion that have been used in medical (and other) selection tests have never been validated against the operational efficiency of an individual or group of individuals doing a particular job in the Navy. For example: Some of the current examination procedures for the Navy require a rather high standard for visual acuity. The standards have been relaxed for certain ratings, or jobs, but why were they imposed in the first place? We know people vary as to their degree of visual acuity. It is also true that personnel need a particular visual proficiency in order to perform certain tasks in the Navy or to process certain information. But, do they have to be as good as the "young, normal male?" Obviously, not all do. We can accept people who have corrections up to a degree. I certainly believe we can tolerate a higher percentage than we do at present.

To take a specific example; for many years there was a stipulation in Naval Aviation that pilots have 20-20 vision. I am not aware of any research that has demonstrated that this "standard" is a requisite for adequate flying performance. It has face validity, but no one has shown the absolute necessity of such standards. Certain tests of depth perception for flyers have been abandoned because it was demonstrated that those who did not have perfect depth perception seemed to get along as well in landing as those who did. We also know that pilots with 20-20 vision, as they get older, usually need glasses for reading for example. The average pilot who has been flying for a period of seven to ten years is getting to the age group where most need some correction. Yet many can "pass" a Snellen chart test. Strangely, some will wear dark glasses almost all the time. Informal surveys have shown that many of these dark glasses are worn in situations in which they are inappropriate for the ambient light conditions.

Some of them, of course, have corrections ground into them.

In an examination of a taxonomy of medical selection tests one gets the impression they were selected on the basis of some judgmental standard of what is "good." and then somehow this score or standard gets into a requirement. Subsequently, the requirement must be met without really knowing if it is apt. Is this level really what the personnel tasks require? Do these capability standards describe the range of abilities truly needed?

Color blindness is another example. For what tasks does a man need trichromatic vision? What is the evidence that he needs to have it? The fundamental question is, not is it good, but are any of his tasks of such a nature that high acuity, wide range color vision is necessary?

Now let me address myself to a subject I know a little more about than some of the other areas, namely, hearing testing. Again, the question arises, "Testing for what?" The regulations, or alternatives to some of the present regulations, state that it is an acceptable procedure for a medical officer, if he has talked to an individual and the man answered his questions, then he can certify the candidate's hearing is functional. For many years the test which was recommended was the so called "whisper test." From a specified distance the medical officer or his assistant would whisper and the examinee was to answer the examiner or repeat what was whispered, his back to the examiner. If he could do this, his hearing was considered acceptable, regardless of what his acuity really was. If he were given a very thorough audiometric examination, he could then, very soon, get out of the Navy and have a service-connected hearing disability. Well, there are whisperers and there are whisperers. Certain examiners with little training can whisper rather loudly. There are medical officers and hospital corpsmen who have developed such skills. This skill was useful for the Navy, particularly during World War II. When it was necessary to obtain great numbers of personnel, often the whisper became quite loud.

Even today there are really no firm audiometric standards. The criteria are in a state of flux, though hopefully changes are in sight. Presently the "normal" or baseline for hearing is being modified. Research has shown that hearing testing in the United States, using the recommendations of the American Standards Association i.e., zero decibel hearing loss, doesn't correspond to the sound pressure level thresholds found for young, normal, male, healthy, adults. The zero reference level is being "lowered" by approximately ten to fifteen decibels at most frequencies. This will become an International Standards Organization prescribed Level for Hearing, and will make quite a bit of difference in viewpoint adaptation for the people who do hearing testing, and for those people who are responsible for the legal aspects of hearing trauma.

Special tests for hearing have been devised for specific situations, particularly, voice tests, speech intelligibility tests, speech reception tests, which are used to assess training progress for telephone talkers, control tower operators, etc. There are no standards for those specific abilities which are Navy-wide. The norms which were obtained for many of these tests utilized the rather poor equipment that was standard Navy issue at that time. Electronic instrumentation has changed in control towers, hence, the precision of articulation that once was so necessary for control tower operators is no longer quite so necessary. For example, it is a rare occasion today when you hear a control tower operator say "fi-eve," as was "standard procedure" in World War II. Such exaggerated pronunciation became standard because of the very poor equipment used to transmit voice messages in those days. This could lead me into a long polemic, that the Navy still would profit by adopting adequate voice communications equipment in all of their communication situations. But, I won't get onto that soapbox.

When a selection candidate comes in, we take his blood pressure. Why do we take it other than for diagnostic purposes prior to and following exposure to certain stressful

situations? There are some interrelationships, reasonably well established, that show if there is a wide differential between diastolic and systolic blood pressure, the chances are increased for physical difficulties to develop at a later date. Once a man is accepted for duty, and unless he goes into the dispensary for a particular reason, blood pressure is rarely taken again routinely. It is well to note that the blood pressure standards were not determined originally by experimentation but were accumulated clinically. In exotic environments, as in space, in SEALAB for example, one of the first requirements is that a blood pressure measurement be taken routinely, every day in SEALAB II. For reaction to certain environments, for fact finding, and for experimental purposes, the test has been useful. However, even here, it is a rare occasion that the measure has been used as a sole basis for changing operational requirements. For example, in all of the Mercury and the Gemini flights, it was taken routinely, and quite often, but it was never found to be related to performance or to result in a lasting medical change.

There are other medical tests that have been used, primarily because someone decided that it was "good" information to have. Moreover, if the man ever became ill, these tests might serve as a baseline against which he could be evaluated later on. In my opinion, this is the major rationale for many of the medical tests that are currently being used in selection batteries. However, I would welcome comments on this point from the Workshop participants who are more sophisticated than I regarding medical testing of this kind.

Let me make a side comment here, again, a personal opinion. I realize it is relatively easy and quite interesting to devise tests evaluating performance and the general medical condition of people subjected to unusual, in some cases, exotic environments. The SEALAB series is a case in point. Test proliferation for this purpose seems to have become almost a fetish with certain laboratories. A great deal of resources are directed into these programs. Apparently the compelling drive of most life scientists who con-

duct such research involving testing in unusual environments is based upon the belief that some of the data may eventually be used for more detailed discriminations among operational personnel. In short, my belief is that one can become possibly too enthralled with the unusual environment testing.

At present, there are a number of diagnostic tests in use. Most agree that there are a number of ways in which many of these tests can be simplified to yield the same information. This is largely the result of improved data collection techniques. The Philco-Ford Chair is an example. The man sits in the chair and has an EKG taken, as well as blood pressure, respiratory rate and body temperature measurement, all this by sitting quietly without any electrode attachments. There are a number of variables affecting the data collected by this instrument. For example, no one has determined how still the individual has to sit. But let me return to my basic theme. Essentially, are medical selection tests useful? For what purpose? Should such a group of physiological tests on an individual be used routinely?

What about an EKG for example? As is well known, electrocardiograms are collected routinely in a number of selection/screening situations. These data are collected from flyers for every annual flight physical. The EKG is probably good as a confirmatory, diagnostic test after there is some indication from clinical observations that an EKG is necessary. A routine EKG doesn't tell you very much except the patterns of electrical potentials of the individual's heart at a particular moment. Some research has been done to examine certain EKG characteristics as predictors of stress responsivity, fatigue, etc., but only with quite limited case samplings. There are some tentative indications that if you monitor the EKG in order to identify changes in the form and pattern of the EKG complex, that these observations may constitute a useful index in stress monitoring situations. Unfortunately, adequate normative data to interpret these findings do not exist for the military.

This leads to another more exotic electrophysiological index — the EEG. I prefer not to initiate argumentation regarding the value of this measure. According to Dr. Enoch Calloway of the Langley-Porter Neuropsychiatric Institute, alpha rhythm seems to correlate only with the cardiac cycle. Moreover, he states that this is the only physiological index with which the unprocessed EEG alpha consistently correlates. It is true the technique can be used as a diagnostic tool once it has been suspected that there are focal lesions or a history of injuries to provide substantial evidence of trauma. Also, if there is a predilection for epilepsy, the patterns usually can be modified by some photic driving technique.

One method presently receiving increased attention as a physiological measure is derived from EEG and is termed Averaged Evoked Responses, (AER). For example, Dr. Carroll White, of NEL, San Diego, using evoked potentials, determined the degree of color perception of individuals over most of the spectrum. The averaged EEG evoked potential picked up from the temporal region will also give a pretty fair indication of the "central registration" of a tonal stimulus. One application of this AER technique is to assess the hearing of people who cannot respond by motor or verbal means during audiometry. The validity of the Evoked Response method of hearing testing is being examined by Dr. Hallowell Davis experimentally at the Central Institute of the Deaf.

Another aspect of the EEG signature recently explored is the "E" wave described most fully by Gray-Walter. There are a number of people who are using this so-called expectancy, or "E" wave as an index useful in monitoring alertness. The pattern is purported to reveal the registration or the perception of the importance of an event for the individual, or the realization that he is supposed to do something. In other words, it is presumed to be indicative of a state of alertness or a state of arousal, though this line of research is still in its experimental stages. The "E" wave has not been used routinely and it is not known precisely the

range of situations in which it might be used.

GSR? By what physical properties are you going to measure it? Impedance? Capacitance? Should one really measure amplitude or amplitude ratio or changes over time? Must the measures be obtained under operant or classical conditioning situations? How does one quantify these factors easily and rapidly? These questions, by the way, are still highly controversial, even among the experts. Finally when you come down to it, what does the GSR measure tell you? With what behavior does it correlate? Many of us have opinions founded on the basis of some evidence, but it would be almost impossible to assess a given situation and state that the GSR should be used under a specific set of conditions or with a specific individual or group.

There are ambiguities even when it comes to the time-tested measure of chronological age. Implicitly, we take age into account in certain selection situations largely because of various empirically-derived statistical relationships found in the literature. It has been observed for example, that within one particular age range people in our culture tend to have more gastro-intestinal disorders. The inference is that there are certain jobs that younger individuals should occupy and be denied to older ones. With age, your spinal column isn't supported with the same muscle tonus that it once had, so one cannot be subjected to an ejection seat experience after a certain age. These practices have emerged from clinically obtained normative data. From such data and in many instances, it is impossible to predict that a particular individual, solely because of his age, is unable to perform a particular task for which he has been trained, e.g., senior aviators. In this context, one task for medical selection would be to determine within narrow confidence limits, the indexes describing the person who could do a job well beyond the age at which the age standards indicate he should be disqualified. Only a beginning has been made in the collection of such data on an experimental basis.

There are a number of other biomedical

indices that may be desirable for a particular environment or for a particular operational situation. For example, there is a test for disorientation that has a great deal of face validity, at least for flyers. But the present indices for disorientation are quite crude. There is a real need for the measure to be more subtle for one thing. Obviously, there are many Navy people exposed to the environmental conditions which yield the syndrome of motion sickness. As is well known, extreme nausea and vomiting are usually accepted as valid symptoms of motion sickness. It is not known how some of the early symptoms, i.e., feelings of malaise, of general discomfort, the perspiration and so on might affect performance before the man reaches a state of nausea and vomiting.

We have observed that man is, in many ways, a remarkable and highly adaptive organism. Careful experimentation is needed to develop a scale providing a third dimension to Dr. Chamber's matrix. This dimension has to do with the degree of adaptation or prior proficiency which can make quite a difference in certain environments as to whether the individual can perform well, or less well, or not at all, e.g., in extreme high or low thermal conditions. This area is replete with measurement difficulties, for example, the problem is still not completely answered as to how to measure "body temperature" and under what constraints (wet or dry bulb?) Where does one place the sensor? All of these factors are important to consider in relating performance or behavior to such measures.

Another series of problems arise in the area of vibration tolerance. Increased vibratory environments are becoming more frequent than less frequent, although the people who build ships and aircraft are now trying to design so as to minimize vibrations. There are not even good simulators to use as test beds nor are there adequate performance criteria to test against.

Another much-used test is the Critical Flicker Fusion, CFF. It has been advanced as a test for many different functions. It is difficult even for those who have spent considerable time experimenting with changes

in critical flicker fusion to explain its behavioral or physiological significance other than as a technique demonstrating an interesting phenomenon and that the processes being measured are probably cortically mediated. Another medical-physiological approach is becoming increasingly important, namely to devise brief, yet valid and reliable tests for toxic reactions to various aerosol environments, to various inoculations and to a wide spectrum of drugs. As many of you are aware, there is a growing proportion of individuals in our population who administer drugs to themselves; drugs defined in a very broad sense. Aspirin is a drug; most of us take it without sanction or without the recommendation of a physician. Also, most of us imbibe rather potent compounds before or after dinner and others more often than that without reference to a prescription. One cannot predict, except by wide and long experience the reaction to this drug. Some people never seem to learn what their tolerance levels are. Tests should be devised for this purpose, to be self-monitored of course.

Another test, and one that has been included in almost every new test battery devised, is reaction time. Reaction time of what, for what, to what? One fact stands out, namely, that a reliable reaction time test devised for a particular application, may have unique limitations and probably should not be applied elsewhere, except with caution. Mathematical techniques are emerging that may be able to help us take the welter of data obtainable from an individual and cluster these in ways so as to provide a composite, usefully predictive of an individual's reaction to particular tasks. In a computer era, this is probably one of the methodological trends that has a high probability to add some sophistication to the selection field.

I have advisedly omitted a class of measures which are designed to "tap" complex perceptual and cognitive functions which, in turn, are interrelated to so-called psychophysiological processes, hence "roughly" in the biomedical field. Figural after-effects might be taken as an example. There probably is much work to be done in this area.

In retrospect, my major point seems to have been a suggestion that we critically examine the reasons for inclusion of our tests in a given battery whether it be for a specific selection task or for research purposes only.

FREE INTERCHANGE FOLLOWING DR. TOLHURST'S OVERVIEW

DR. GUNDERSON: I'd like to mention one kind of data we are collecting in San Diego. In Underwater Demolition Training, as you probably know, men are subjected to very rigorous and arduous conditions involving loss of sleep and other very severe circumstances. We were interested in whether we could predict certain biochemical changes as well as attitudinal changes during the course of this training. We also were trying to predict attrition—not only in the sense of identifying those who were dropped from the course, but when, during the training, they dropped (early, late) and under what circumstances. One aspect of the study involved taking blood samples periodically during the training period, the analysis aimed at determining the uric acid and cholesterol levels. We were very surprised when we found that the uric acid levels were extremely high for all subjects tested, far above the averages for the general population. But as training progressed and as the men began working harder, physically, in their training program, uric acid levels dropped sharply, levelling off at a normal level. This finding may seem opposite to what might be expected. We interpreted these data as indicative of the concomitants of anticipatory anxiety, resulting presumably from something like "fear of the unknown." We've replicated this on another class and it was interesting to see that this kind of relationship holds up. It is also interesting in relation to the sick call data collected aboard ship, data mentioned earlier in the Workshop in another context. We interpret the relatively high incidence of sick call going into combat as compared to the rate in actual combat also as symptoms of anticipatory anxiety.

DR. RADLOFF: Was this uric acid level related to individual differences in performance?

DR. GUNDERSON: We have not had enough subjects to do anything with that question. The striking thing was a uniformly high level for all subjects. We know that these men are in many ways unique in terms of motivation. They are extremely involved with this program as a rule; they are striving individuals. We think this unique motivation is systematically related to certain personality variables. Yet, we haven't been able to relate the differences in levels to attrition and so on. One possible use of data of this kind might be to identify potential early dropouts from the program. When is a man showing extreme signs of physical and biochemical stress? We hope to be able to relate the blood data taken from our second class, which contained more subjects, to this particular criterion. We also have behavioral data and subjective symptom data (moods and the like) to be interrelated to the blood measures. Our procedure, by the way, was to take initial baseline measures when they came in to UDT. Even on the first day when they thought they were getting the pretesting, we actually got extremely high blood uric acid levels.

DR. SCHAEFER: During an isolation study, we got elevated uric acid before, as well as during, the study. These findings seem consistent.

DR. GUNDERSON: I might add that the cholesterol levels showed opposite changes, down initially, but with some elevation during training.

MISS AMBLER: A so-called "Vestibular Disorientation Test" has been developed at Pensacola. It involves the assessment of reactions to a mild vestibular stimulus. We have gotten some validity against a criterion of anxiety in flight training. Moreover, we have gotten cross-validation as a matter of fact. We feel this may be an important finding.

One general comment should be made about biomedical indices in pilot selection, blood pressure for example. Measures of

this kind would appear to have relevance mainly in the sense of providing base-line data for future reference in clinical situations. Longitudinal "health status" studies involving measures of this kind among others appear to be useful. The Pensacola "1000 Aviator" Study is a case in point.

DR. HUNT: Haven't we "run into" the same type of problem we ran into in the previous Workshop Session on personality tests? If your biomedical tests are specific enough and used to predict specific activities, we may have something. But do we have trustworthy normative or as it has been called, base-line data against which to evaluate our findings?

I believe our medical colleagues would agree that existing normative data for most medical measures are less stable than desirable. Often inadequate training for test administrators is at fault. Too, the people who interpret the data may be poorly trained and lacking in breadth of experience. I certainly think it is fair to make these statements about the general medical examination given in Navy dispensaries and recruit centers.

DR. MOONAN: One kind of question a methodologist might raise regarding biomedical data has to do with changes in many of these scores over time. For example, change with age, I would judge, affects initial selection criterion levels with respect to variables of this kind.

It has been mentioned or implied during this Workshop, that there is some relationship between biomedical indices and personality dimensions usually "tapped" by personality tests. I have a suggestion, again methodological, that predictive relationships are often "strengthened" by the application of stringent item analytical rather test score validation techniques with respect to a given criterion, say a biomedical one like blood pressure.

LT GREEN: The point has been made, several times during this Workshop Session that we try to select people whose condition, as assessed by biomedical and other indices, does not deteriorate rapidly with age. At present, we are forced to consider the Group 4 aptitude problem. Similarly, we may even-

tually have to consider for recruitment men with biological characteristics that are not so desirable. I think we need to re-examine some of our biomedical testing programs with this fact in mind. Most certainly, in biomedical measurement as in all measurement, we should have a "solid" rationale underpinning the measurement technique. The so-called "shotgun" approach should be avoided.

DR. CHAMBERS: Beyond a doubt, the innovations that have resulted from modern electronic circuitry and from the introduction of computer techniques are many and varied in this field. Take as one example, the telemetering field. We have transmitted biomedical information from Houston to Rutgers where it was analyzed and returned so that the medical personnel could study the results as they occurred.

Despite all these advances, it seems to me that we have a real problem, a problem alluded to by Dr. Hunt a number of times during this Workshop. Data of this kind, in certain instances, seems to take on a very personal meaning, and as a result, we are unable frequently to make it available to the people who might be able to benefit most from the information. Often it is relegated to a medical folder which may be "off limits" to the researcher. This may or may not be good, but, in any event, the problem of its utilization for purposes of selection raises some very serious questions as to how it should most judiciously be handled. For example, pilots and operators of complex equipment in the Navy, in NASA and in the Air Force characteristically have two types of medical folders. One goes with them wherever they go. This is the folder that has the records that are available to their own command. Then there is another folder, usually, that has all this special information, some of which is often useful in our research, for example, the results of special biomedical tests such as tests of oxygen tolerance, electroencephalograms and other special test results that are often unavailable for anyone to use even for research purposes. In fact, there are instances in which identification data are removed from the folder so that it can't be traced back to see which man this particular set of data belongs to. This is a

matter of protecting the individual's personal medical data which often gets treated quite differently from other types of information.

DR. GELL: I'd like to make a case for blood pressure measurement which appears to have been downgraded in terms of its usefulness. I think it is one of the best indicators we have of states of agitation or as the case may be, states of complacency.

As you know, blood pressure is one of the most important indicators on the lie detector polygraph tracing. Almost thirty years ago, we conducted a study at Randolph Field in which blood pressure responses to ice water emersion were measured on about 275 flight candidates. Individual differences were remarkable. For example, three men showed no response at all. On the other hand, three or four showed extreme changes and extreme lability of blood pressure. The majority, of course, were in the middle. All in all, the relationships between these blood pressure change-measures and other measures of emotionality were not strong. Of course, modern techniques of indirect blood pressure measurement are much improved over the methods available in those days, so that it might be well to repeat some of the older studies involving this and other biomedical indices in common usage.

DR. CHAMBERS: Certainly instrumentation and general techniques are very important if measures of this kind are to be of maximum value. I have recently read a report from the Bureau of Standards, which conducted various tests on a variety of polygraph instruments. Apparently, there are great differences between instruments in terms of their linearity, ease of calibration and so on.

DR. SCHAEFER: I would like to make a few comments interrelating two of the Workshop Sessions, the one on Personality Measurement and this one on Biomedical Indices. Apparently, one conclusion from the personality session was that personality tests are dead. In the second one, unless I am misinterpreting the comments biomedical indices are not so good, to say the least.

Let me suggest a different methodology to apply to the biomedical measurement prob-

lem. Our work suggests that the physiological cycles of these functions taken minute by minute are important indices of underlying psychological and physiological processes. Note please, I am suggesting instead of relying wholly on static, basal or base line measurement of these biomedical indicators that you focus upon dynamic, often rhythmic, (sometimes circadian) fluctuations of these measures over time. We have some data in the literature that argue for the interrelations between these change-measures of physiological functions and different personality types. Though our subject sampling is sparse, nevertheless, the data are suggestive.

One might hazard a guess that interpersonal compatibility in certain situations, confinement for example, is somehow related to the degree to which the person's periodicities of physiological functions are synchronized. Preflight training data from two Russian astronauts certainly argue for this possibility as do the data from our studies of diurnal rhythms.

In my opinion, Dr. Tolhurst gave a good summary of the biomedical indices, but I think it was rather negative as a whole. Negativism may be justified with respect to some of these tests, particularly when they have not been studied experimentally. Too, there is the possibility that the negativism has transferred from the previous Workshop Session dealing with the use of Personality Tests in selection.

However, let me make a statement in defense of biomedical tests generally as selection techniques. To repeat, I believe Dr. Tolhurst's excellent overview of this area, while comprehensive, seemed slightly negative in the sense that many of these measures are taken without a clear rationale for their use spelled out. The requirement that the reasons for taking data of any kind should be thoroughly developed in the context of what is known about the various indices used and more importantly, the processes these indices are designed to measure. As is the case for all measures, when the question of what use is to be made of the measure, a validation question, we need to extend the question

"valid for what purposes?" For example, consistently, high diastolic blood pressure has validity for predicting kidney malfunction of one kind or another. So, maybe we ask the wrong validation question for some of these so-called biomedical measures. Also, in an age of ever accelerating sophistication in instrumentation technology, one needs to exploit every advance in order to get maximum measurement reliability. This takes money, a lot of money. For example, to collect complete pulmonary function test data in order to estimate the work capacity of astronauts, someone has estimated the cost to be about \$1,500 per man. But the expenditure is necessary if you want to collect meaningful data.

DR. TOLHURST: Dr. Schaefer, possibly this perceived negativism on my part comes from a misunderstanding. One of the main points I tried to make was that normative data of acceptable accuracy for many of these biomedical indices, particularly as related to performance in a given task, are lacking. As a result, we have few trustworthy standards or criteria against which to interpret our data. However, I did not mean to imply that biomedical tests of various kinds taken in a clinical setting, often for diagnostic purposes, are not useful. Indeed, they are necessary as indicators of general health but when considered predictors in a selection program, one needs to be quite cautious.

DR. WHERRY: As to periodicity of these measures, our studies at Pensacola showed that changes from a man's base line for some of these measures are predictive of deterioration later on. But individual differences within a given age group are so great that it is very difficult, if not impossible, to separate a normal healthy individual from an unhealthy one, or more importantly perhaps, to be able to identify who will become unhealthy. It may be possible to get some help from certain personality test data. Throughout a man's life span, he develops skills, habits, coping mechanisms someone called them, to meet the demands of his environment. If the person falls in an acceptable range (often poorly delineated), we can presume these mechanisms were adequate. As an example, person-

ologists have in the past, spoken negatively of the so-called "rigid" personality. I wonder, would we have selected a presidential candidate so rigidly honest with a high squeaky voice and with the other attributes that Abraham Lincoln had?

I would like to make some additional comments about the anticipatory threat concept mentioned in the previous Workshop Session on Personality Testing. Let me first state categorically that I am in favor of biomedical testing in certain selection programs. Data are coming in to argue that the situational testing concept wherein measures of this kind are taken can be used to predict behavior in real life (for us, military) situations. In short, there seem to be some useful relationships between these laboratory data and individual differences in a man's behavior when faced with threats of various kinds out "in the field."

Let me mention an experiment that was done at Pensacola by cardiologists during 1, 2, 3, and 4G maneuvers. One cardiologist I recall was excited by some inversions of "T" waves observed during the exposure. The observation took on new meaning, however, when it was discovered that some of these inversions of "T" waves were happening before the men started the high G exposure. Something like anticipatory threat seemed to be involved. Other measures such as blood pressure and respiratory and heart rate also showed similar changes just prior to the actual exposure.

I realize there is a further question, namely, do the patterns of responses to threat relate in any systematic way to quality of performance carried out under the same conditions? I don't assume, and have little data to support the prediction that a man who shows significant physiological changes in a threatening situation will show performance changes (increments or decrements) of any magnitude. But is it possible to demonstrate that the physiological changes are associated with performance degradation in the field, say when coming on duty on an aircraft carrier, or just before making an ascent in the submarine escape training tank are of the same

kind but of different magnitude as those observed in the laboratory? If they were, certain adjustment predictions might be possible.

DR. TOLHURST: I wouldn't be in the position I am in in ONR if I didn't believe that there were some meaningful relationships between behavior and physiology.

DR. GUNDERSON: Let me re-emphasize in this context, the logistics problem that Dr. Chambers mentioned earlier in this Workshop Session. The problem is that the general health data are in the man's health record; this is highly useful information about the man. But, in addition there are files of data pertaining to his reactions to the centrifuge, to the training tanks or some specific biomedical test. These latter data mentioned are crucial, but often the two kinds of information on a man are not available at the same time and in the same place.

LCDR GALLAGHER: Two uses of biomedical data seem to have been confused from time to time during this session of the Workshop; that is biomedical indicators used in screening programs and biomedical monitoring during exposure to unusual environments. Take as an example the Snellen Eye Chart data for predicting visual efficiency in an aircraft cockpit. First of all, few, if any, tasks in the cockpit require 20/20 vision. Incidentally, I don't believe we have substantial data pertaining to visual acuity changes as a function of aging. Longitudinal data for hearing and other biomedical functions are similarly lacking. These data are needed for various populations. The Pensacola 1000 aviator study may be a start in the right direction.

So we need to set biomedical standards contingent upon the kind of duty the man is assigned. Take auditory standards for another example; it would seem to be a fair assumption that anyone who has been in aviation for ten years has a hearing loss. However, it is well known that a pilot can get along in certain types of aircraft with a 10 to 20 db or more hearing loss. The same may not be true for a sonarman however.

*Dr. Polis is presently at the Naval Air Development Center, Johnsville, Pa. His recent work with pilots is to be published in the *Journal of Aerospace Medicine* in 1968.

Let me make a few comments about biochemical measurement. In one recent study by Dr. B. David Polis, blood phospholipid analyses were made on men on high risk missions over North Vietnam. Differences in this and other biochemical measures were seen in the pilots just before the end of their 100 day high risk duty assignment as compared to a control population. Differences were also found for hospitalized schizophrenic patients and for normal groups after sleep deprivation and after high "G" exposure. I want to emphasize that the changes in the biochemical indices obtained from the pilots were not associated with performance changes of any significant magnitude.*

DR. WEYBREW: In my opinion, Dr. Tolhurst made some excellent points in the context of his overview of a very complicated field. Over and over, it seemed to me he was admonishing us to ask ourselves a fair question, simply put, "what is the rationale for using this or that measure in this or that situation?" In this context, I am reminded of Colonel Simons' answer when asked, "why were they taking blood pressure, and why were they monitoring this and that during his balloon flight?" He answered something like, "because we had the equipment to do the measurement." I say this in all sincerity—the drunkard's search paradox is still with us.

Speaking first to a point that Dr. Wherry makes in connection with his stress studies, I think one needs to look at physiological indices in the context of the situations in which the data were collected. I think it is interesting and within limits, informative, to take measurements during "shock threat," but I think one needs to look at the correlation between responses to shock threat in laboratory situations and similar responses to more "real-life" threats. My experience leads me to the prediction that the correlations will be anything but remarkable.

Another general point is that the physiological functions that I think we should be interested in are dynamic and not static. The most important variance of these measures may be the intra-person variation over time rather than the variance between persons at

a given time. This would seem to be so particularly when attempting to understand a given person's stress adjustment from situation to situation.

Perhaps I am behind the times, but periodically my medical colleagues and I have talked about autonomic nervous system function. I have had medical people tell me that autonomic system function is generally covered somewhat cursorily in courses in anatomy and physiology. The main point is that there are any number of books on autonomic function, particularly as related to behavior. But, unless my observations are off, most of this writing is overwhelmingly coming from psychologists or psychophysiolgists as they have come to be called. I don't seem to run across much on the subject in the physiological literature. I'm indicating a bias I have and am recommending some interdisciplinary collaboration to try to come up with some new ideas in a somewhat poorly delineated field.

I believe that some of the biomedical indices of the physiological tone of a person are predictively useful. Dr. Ax* has proposed a concept that may have something for us. He calls it Physiological Learning Aptitude, which is related to how conditionable the autonomic nervous system is.

Another related idea which may prove to be useful is Gellhorn's concept of Autonomic Tuning.** In an excited condition, you may get a blood pressure rise. In a less excited condition the blood pressure may drop to the same situation depending on the pressure level before stimulation.

There is nothing new about the finding that, for some measures, pulse and blood pressure for example, there characteristically is a correlation between change and level. In fact, as I understand Wilder's Law, it is concerned with this matter. But the notion of looking at this level/change relationship as a parameter with predictive potential is new, at least from my point of view.

DR. WEYBREW: I think something like Malmö's Activation Level Concept is useful.* I heard Dr. John Lacey, sometime ago** take five minutes to sound the death knell for the term and anything like it. I frankly think we need some concept like this. It is like the term attitude, in a sense. The question of a behavioral referent for the term disturbs some people. We need some term like "activation." I can point to behavior that reflects an attitude. Similarly, I can point to different levels of activation or excitation of people. I know when I'm steamed up and I can, or think I can, observe excitement or tension (activation) in others. Therefore, I think some concept like activation or excitation level is important and, potentially at least, useful in selection.

DR. TOLHURST: One generalization stemming from this session overview and the free interchange following might be that there is some fresh thinking badly needed in an area somewhat lacking in guidelines for the directions that thinking should take.

*Ax, A. F., Beckett, P. G. S., Fretz, N. A. & Gottlieb, J. S. Development of a selection test for motivational aptitude. Detroit, Mich.: Lafayette Clinic, Contract No. NAS 2-1031, 1965.

**Gellhorn, E., *Autonomic imbalance and the hypothalamus*. Minneapolis: University of Minn., 1957.

*Malmö, R. B., Activation: A neuropsychological dimension. *Psychol. Rev.*, 1966 pp. 367-386.

**Lacey, J. I., Somatic response patterning and stress. Some revisions of activation theory. In Appley, M. H. & Trumbull, R. (Eds.) *Psychological Stress*. New York, Appleton-Century-Crofts, 1967.

SESSION V

SMALL GROUPS OBSERVATIONAL DATA IN SELECTION

SESSION OVERVIEW SMALL GROUPS OBSERVATIONAL DATA IN SELECTION

William Haythorn, Ph.D
Naval Medical Research Institute
National Naval Medical Center
Bethesda, Maryland

Some of the problems that seem to me apparent in trying to consider small group observational data for selection purposes might be illustrated by the behavioral function formula developed by Kurt Lewin, to the effect that behavior is equal to a function of the person and the environment interacting ($B = f(P, E)$). The conference would appear to me to have shown so far (in our discussion of criterion measures) that we don't know how to measure behavior (B), particularly as it relates to performance effectiveness and adjustment criteria. Our discussion of personality variables indicated that we also don't know very much about measuring the person (P), either from the personality or physiological point of view. Moreover, we don't know what to do with such measures when we get them. We will turn now to completing the equation by discovering that we don't know much about measuring the environment (E), either. When we talk about situational tests and of generalizing from one situation to another, we are as limited as when talking about the other two terms of this formula.

There are at least two possible interpretations of the question, "What value do small group observational data have in selection?" One of these would be: What is the value of situational tests? and the other: What is the value of collecting data on small groups in real life situations and using the data for prediction purposes?

Our earlier discussion said a bit about the latter—namely, that in a real life situation like SEALAB, we can collect some very useful observational data if we are willing to pay the cost. In a training situation, in recruit training, or in submarine training, we get a great deal of information of an observational nature which is, or may be, used to screen out certain people. We haven't however, said very much about situational tests.

I might start by saying that I don't have a great deal of confidence in situational tests for selection purposes at this point in our state of knowledge. One of the many problems in trying to use situational tests or many other kinds of observational data for selection or prediction is that we frequently make assumptions of linearity of the variable relationships. This reflects the limitations of existing models. Though, I understand statisticians are making progress toward solving this problem by means of a procedure for using multiple regression in non-linear relationships. Nonetheless, in a typical selection research program linear assumptions are predominant, and my feeling is that linear assumptions may not adequately fit reality. It might be profitable to look for other kinds of relationships such as those involving non-linear contingency or interactive effects. In dealing with personality variables, which frequently become involved in small group observations, any particular personality variable relates to any criterion measure usually only as a function of the particular values of other variables. So one would have to take into account contingency relationships. This goes beyond just dealing with personality measures in individuals. A more sophisticated approach is illustrated by a crew composition study done by Bill Voiers at the Crew Research Laboratory (CRL), Randolph Field, many years ago.*

Studies at CRL of B-29 crews as they came through the pipe line on their way to combat in Korea consistently showed that

*Voiers, W. D. Bombing accuracy as a function of the ground school proficiency structure of the B-29 bomb team. Lackland Air Force Base, Texas: Air Force Personnel and Training Research Center. Jan. 1956. (Research Report AFPTRC-TN-56-4).

individual proficiency measures, for example, individual ground school grades and individual flight training grades had practically zero correlation with crew effectiveness, as measured by simulated bomb drop scores. Part of this was because the criterion measure was highly unreliable. That is to say that simulated bomb drop scores on one run didn't predict performance on the next run. They did have some reliability, however, although we couldn't predict any of the reliable variance from individual crew member scores. Voiers began looking at the situation as an interactional relationship among members of the crew and by correlating, for example, the proficiency of the radar bombardier with the team score, he found positive correlations only if he limited the analysis to teams that had highly proficient aircraft commanders—i.e., pilots. If the crew had a relatively poor pilot, it didn't make any difference how good the bombardier was individually; he couldn't significantly affect the team scores. Similarly, the proficiency of the pilot had no relationship to crew effectiveness except when the crew had a better than average bombardier; it didn't really matter how well the pilot flew the plane approaching the target. If the bombardier couldn't do his job well, the effect of pilot proficiency on the team score was essentially zero. This is part of the background of my own interest in looking at compositional variables, including not only interaction effects among the variables describing single individuals, but also interaction effects among scores describing different members of a team.

The prediction of crew performance from individual member characteristics is yet more complicated. I don't believe it is very useful or promising to correlate directly from an individual score on any test to a criterion measure of any kind except in very limited circumstances. A crude conceptual model that we have been using at Bethesda deals with three classes of independent vari-

ables: (1) composition factors describing individual differences and team clusters of individuals; (2) environmental variables; and (3) management intervention variables. (By the latter is meant what the organization does to the men—training, leadership, incentive programs, and so on.) The model envisions these three classes of variables interacting to produce intra-group processes. Team effectiveness, as opposed to individual effectiveness, results from these group processes. What this suggests, to me at least, is that instead of the usual practice of relating independent variables to criterion measures, we need a better description of how the criteria are affected by intermediate processes such as communications effectiveness, motivation, morale, and the like. These in turn can be related more directly to the traditional independent variables.

The emphasis on small group observational data results primarily from the fact that social skills and other social variables relevant to individual behavior cannot be measured in any other way except from observations of the man in a social situation. One turns to situational tests or observational data generally when he is interested in leadership, ability to relate to other people, social adjustment, and other concepts of that nature as opposed to psychomotor performance or tasks involving specific knowledge. The latter processes can be measured more adequately in other ways. It seems therefore, that when we consider small group observational data for selection, we are concerned with selection for situations where social skills and interpersonal relations are relevant.

Now, in what situations are they relevant? Again the Randolph Field experience is basic to my own view of things. We were studying B-29 crews as they came through Randolph and went on to Korea.* We collected all kinds of data, as psychologists often do under these circumstances, and tried to identify variables that might predict performance in combat. We obtained aptitude scores, personality scores, performance measures of the team in training, and attitude scale responses. We

*DeGough, R. A. and Knoell, Dorothy M. Attitudes relevant to bomber crew performance in combat. Lackland Air Force Base, Texas: Air Force Personnel and Training Center, May 1954. (Research Bulletin AFPTRC-TR-54-18).

found, much to our surprise, that performance in training didn't predict performance in combat at all. There was essentially a zero correlation between training performance and performance in combat (as judged mainly by supervisory ratings). However, attitude scales administered in training did significantly predict combat performance. One set of results worth mentioning very briefly involved two of our attitude scales, crew cohesiveness and crew motivation. Crew cohesiveness was defined as responses to items asking how well they liked other members of the crew, how much they wanted to stick together after the war, how much they went on liberty together, and things of that nature. Crew motivation involved items describing the crew as trying to do its job well, convictions that the Air Force mission was an important one, and things of that nature. By classifying crews as highly motivated and poorly motivated on the basis of average crew attitudes and doing likewise with cohesiveness, a 2 x 2 matrix of crews was generated. We then entered into this matrix the percentage of scheduled missions that were aborted. The results were approximately as follows: in high cohesive, high motivated crews, 4% aborts; under high cohesiveness, low motivation, 8% aborts; low cohesiveness, low motivation, 9% aborts; and low cohesiveness, high motivation, 17% aborts. The difference between high and low motivation is statistically significant, and the interaction term is highly significant.

The relevance of these results to this workshop session would seem to be that social psychological variables, that is, attitudes measured in training, predicted a "hard" criterion of performance in combat several weeks or months later, whereas objective performance measures in training didn't predict the same criterion at all. This would seem to raise some questions for selection that deserve further consideration.

To use observations of small group behav-

ior for prediction purposes, they must be dimensionalized in some way. There have been a number of attempts to narrow the number of dimensions of small group behavior, primarily by means of factor analytical techniques. The results of these studies seem to have produced some consistent trends. There have been a large number of such analyses, for example, the work of Schutz* and Bass**, which generally yield three variables (factors) more or less consistent from one set of data to another. Generally, there emerges a set of behaviors or a factor that can be described as self oriented behavior, or striving for individual prominence, or seeking control, or dominance oriented, or some such variable as that. The trait or dimension is apparently related to prominence in the group, his capacity for controlling or dominating the group, and so on. A second trait cluster or factor has been variously described as interaction oriented, seeking group approval, or a need for affiliation or affection, a set of co-varying measures that suggest orientation towards other people as a major motivation dimension. A third commonly-found cluster is variously called task orientation, striving for goal accomplishment or need for achievement. The consistency in these analyses, based as they are on different sets of data obtained from different sets of subjects by different investigators, justifies some hope that it may be possible to quantify what is observed in groups in ways that will give fewer dimensions than the almost infinite number that psychologists seem to dream up when they start making their observations.

It appears then, that observations of small group situations, whether they take the form of peer or observer or supervisor ratings, will tend to cluster in these three groupings. The question for selection is: What value are measures of these three kinds of behavior in predicting what an individual will do in a given criterion situation?

Let us turn to the question of situational tests. We have the same problem with situational tests that we used to have with measures of ability. Ability turned out, it now seems, to be multi-dimensional, yet we

*Schutz, W. C. **FIRO: A three-dimensional theory of interpersonal behavior.** New York: Rinehart, 1958.

Bass, B. M. **Leadership, psychology and organizational behavior. New York: Harper, 1960.

seldom hear arguments leading to the conclusion that there are different aptitudes and that these aptitudes are differentially predictive of behavior in different situations. The same thing, I believe, is true in social situations. These also are multi-dimensional, although we know very little about what these dimensions are. Moreover, we are unable to describe the similarity between one kind of social situation and another in any effective way. Yet, if we are trying to predict from a situational test of any kind, (say, leadership, social skills or whatever) to a real life situation, it would be unrealistic to expect to be able to do that unless we can describe the similarity between the two situations. As of now, we can't do that. This accounts partly for why we can't predict very well from training situations to field operations. The similarity between the two situations is usually an unknown quantity. Unless we can assure ourselves that they are similar, there is no reason to expect high predictability.

Part of our program at Bethesda involving isolated groups of men may help illustrate some of these points. It is not a selection study in the usual sense of getting measures on individuals and predicting what they will do in a field setting. Rather, it involves a fairly complicated laboratory setting in which pairs of men were placed in isolation, deprived of contact with the outside, and subjected to systematic manipulation of the environment (including crew composition in one study). In one study we have just finished, men under conditions of isolation gave responses to adjective check lists that indicate a moderate elevation of subjective stress, partly as a function of the composition of the 2-man groups. Hypothetically incompatible pairs of men are likely to give us this elevation more than hypothetically compatible pairs. On three of the four personality dimensions manipulated in the study, the three that yielded significant composition effects were need for dominance, need for achievement, and dogmatism. Need affiliation was also included, but didn't show the incompatibility-stress effect.

Task performance, though, complicates our prediction problem. We had hypothesized

ed that the isolated situation would elevate stress, which it did, but we also expected this stress would interfere with performance. That was not the case. As a matter of fact, there was an enhancement of performance effectiveness under conditions of elevated stress. This finding has been reported frequently in the literature, and is taken to mean that the relationship between stress and performance is a curvilinear one with elevations of stress up to some moderate level improving performance, beyond which there is a performance decrement.

These results have potentially important implications for selection research. It seems to me that most research concerned with group composition assumes that composing compatible crews would be most desirable. Crew members would presumably get along well, they would be happy, they would adjust interpersonally, and everything would be great. These data suggest that such an approach could be predictive of intra-crew harmony, but performance effectiveness might be another question. The compatible crews aren't necessarily the best performing crews and, in fact, in our limited situation, the moderately incompatible crews performed most effectively.

Let me refer again to the B-29 crew study of cohesiveness and motivation mentioned earlier as an illustration of this complex situation. Here were two kinds of compatible crews: crews that were compatible and motivated had very few aborted missions; while crews that were compatible but thought that they were wasting their time in Korea performed considerably less well as measured by the criterion of percent of scheduled missions aborted.

By way of summary, the prediction model we have seems to me to be an inadequate model. There are many "slips" between the individual's capabilities and personality and the crew's effectiveness. Among these "slips" are the subjective stress that the individual feels; the particular nature of the situation in which he finds himself and which may differ from one crew to another or from one man to another; and the particular combination of other people with whom he deals. All

of these should be incorporated into our model. To the best of my knowledge, this has not yet been done.

The protagonist's arguments for situational testing can be listed under a limited number of headings. One is that it is more representative of the behavior to be predicted. If one is concerned with leadership behavior and wants to predict a man's leadership ability, I don't know of any pencil and paper test that is very useful. Fiedler has one (LPC, least-preferred co-worker)* that recently has begun to show promise, but that's the only one I know of. Presumably, behavior in a situational test is less subject to falsification. That's an assumption for which an argument could be made, but I'm not sure that it's true. However, situational tests may be the only way to assess certain aspects of an individual's social or interpersonal capabilities.

The criticisms are that the similarities in the situational test and the field conditions to be predicted are unknown. In fact, we don't even, as yet, have a way of describing the two situations. Group characteristics are highly variable, and being able to measure what a man does in one group setting or laboratory situation may not give very much ability to predict what he is going to do in a field social situation, quite possibly different in a number of ways. Situational tests have not been standardized to any noticeable degree. Bass has done some effective work here, but it is not very extensive as yet.*

Groups change over time. Data obtained on a group of men at one point of time may be quite different from those obtained at a later time. Tuckman recently reviewed the literature on group development and identified four stages that he called forming, storming, norming and performing.** The "forming" is when the group is just getting together, getting to know each other. "Storming" is when they begin to clash over ideals and ways of doing things, and begin adjusting to each other. Groups that

become effective usually get through this "storming" period and begin to develop shared norms of behavior, after which they begin to perform effectively. If data from a group is obtained during the "storming" period, it may not have any value in predicting what he will do in the "performing" period. Thus, the temporal changes in group development also complicate the prediction problem.

Turning now to directions in which future research might go, I think that we need more emphasis on the measurement of social skills. I don't know of a single inventory of what social skills there are to be measured. A lot of people talk about empathy, leadership ability and sensitivity and a lot of other concepts of that nature, but I don't know of anyone who has compiled an inventory of relevant social skills. There ought to be some identification of what we mean, what social skills we are talking about, and how we measure them. I have some hope that 20 or 30 years from now we'll know enough about group composition to begin "grinding" that into selection procedures. I don't believe we know enough now, but I think that, particularly in predicting the relationship between personality and performance, one has to take into account the nature of the social situation, including the characteristics of other group members. If we can develop the ability to select men to form a team, I believe we may have a better chance of predicting effectively than if we select men on the basis of their individual characteristics alone.

FREE INTERCHANGE FOLLOWING DR. HAYTHORN'S OVERVIEW

DR. WEYBREW: I think you are right in pointing out the importance of social skills in group adjustment. The problem is to get a "hold" on these concepts operationally. Some thirty years ago, a group at Minnesota tried to develop a paper-and-pencil test to "tap" social attitudes presumably related to

*Fiedler, F. The contingency model: A theory of leadership effectiveness. In H. Proshansky and B. Seidenberg, *Basic studies in social psychology*. New York: Holt, Rinehart and Winston, 1965.

*Bass, B. M. *Leadership, psychology and organizational behavior*. New York: Rinehart, 1958.

**Tuckman, B. W. Developmental sequence in small groups. *Psych. Bull.* 1965, 63, (No. 6) pp. 384-399.

social skills.* Their approach hasn't helped us in submarines very much. Too, Harvard's Professor Tagiuri has, I think, an interesting approach to group processes analyses in his so-called "relational analysis" approach. It might be worth looking into. Incidentally, a considerable amount of the data from which his theories of social perception emerged was provided by enlisted submariners.**

Another related class of data pertaining to group processes and important, I feel, for understanding submarine crews, is the question of attitude change. Many disagree, but I still believe there is something to the concept of "attitude" as a polarized motive which can be measured. Moreover, if you can "pin down" the factors, conditions, situations, or manipulations that affect attitudes, you can predict attitude change. This information should be related clearly to military retention, if one is looking for a worthwhile application. As an aside, group therapists need this kind of information too. Indeed, attitude information of this kind should be helpful to us in selecting submarine crews.

In the early sixties,* we did a study involving one crew of 125 men submerged on a Fleet Ballistic Missile Submarine for sixty days. We measured attitudes before they submerged and when they came up sixty days later. As compared to a matched control group (the "off" crew for the same submarine), the data argued strongly that interpersonal attitudes definitely change under these conditions. We thought that by using these attitude change measures as a criterion, we could predict individual differences in attitude change from our selection test scores and by so doing, could predict who would

volunteer to go on the second submerged cruise. I am sorry to report that attitude changes as we measured them are correlated with absolutely nothing. We partialled initial attitude level out from attitude change in one study, but still no relationships turned up.

The problem may have been with attitude item content (attitudes toward the Navy in general, towards authority, towards war deterrence concept, attitudes indicating inter-crew member confidence and the like). The idea may still warrant some looking into, however.

Then, there is a third concept. I recall an approximate biblical quote something like, "When I became a man, I gave up childish things." I think the characteristic called maturation, or some concept like "social maturation," might have some relevance for predicting group adjustment. Freud put his finger on the pleasure principle which, as I understand it, identifies the impulsive, acting out (immature?) traits often found in the age groups we work with in submarines. These kinds of personality traits show up in the inter-relationships between persons within the subgroups making up a submarine crew. To get at this trait of immaturity operationally, I think we will probably have to do it by looking at the subjects in some group problem solving situation. We have precious little data that bears even remotely on this matter. As a start some years ago, we took John R. P. French's ball-and-spiral task and put random samples of four enlisted men on it to see if the impulsive, immature traits would stand out during group performance in a sealed submarine situation. Although we never completed a "right-tight" study, nevertheless, from our observations, I felt then and still do, that the approach has possibilities for assessment of team performance.

DR. TOLHURST: Somewhat in answer to Dr. Weybrew's comments about what he called social immaturity, I know of a scale which might get at the trait of childishness or acting out kinds of behavior. It has been

*Williamson, E. G. and Darley, J. A. Minnesota inventory of social attitudes, New York: Psych. Corp., 1937.

**Tagiuri, R. Social preference and its perception. In R. Tagiuri and L. Petrullo (eds.) *Person perception and interpersonal behavior*, Stanford, Calif. Stanford University Press, 1958 pp 316-336.

*Weybrew, B. B., Molish, H. B. and Ninow, E. H. Attitude changes during and following prolonged periods of marine submergence. USN SubMedRes Lab Rpt. 369, 1961. (Contents confidential)

called an impulsivity scale.* It does not correlate at all with many other types of scales, particularly one that is thought to be the antithesis of the trait, namely, anxiety.

DR. WEYBREW: I thought you were going to tell me about the so-called Machiavellian Scale, shortened to the Mach Scale.** I understand this scale gets at tendencies for social manipulation, deceit, unscrupulousness and the like.

DR. RIMLAND: I have a suggestion having to do with combining people into teams. I am impressed—or depressed—with the fact that most of the personality tests that have been tried and have failed, for the most part, have depended on hypothetical constructs of some sort or another. Dr. Haythorn remarked about the several scales that a number of factor analyses have consistently yielded: self-orientation, task orientation, need for affection, and so forth. Even when such scales have emerged from several factor analyses, they still appear to be rather obscure and fuzzy. It is interesting, but possibly not too informative when factor analysis tells us that there is something there, but the relationships are not really very strong, and the scales are really not very useful. To me, it seems that a better approach for dealing with problems of interpersonal compatibility might be derived empirically from specific item content, rather than from any vague hypothetical construct. I became interested in this approach when I was asked several years ago to develop an experimental marriage compatibility scale. The idea occurred to me (and I later learned that it occurred to Lewis Terman thirty years before I thought of it) that if one wants to find a way to pair people, one should try to match those who are interested in and tend to give the same responses to the same stimuli. If I have a single fellow who is interested in water skiing, camping, Shakespeare and folk music similar to the interests of an eligible female, I would consider trying

to introduce them to each other. What one ought to try to do in assembling crews or teams, perhaps is try to find people who have the same interests. I would guess you would get more cohesiveness if the crew members share somewhat the same biases, values and interests. It may be that water skiing, camping, Shakespeare, and folk music define no underlying dimension. I don't particularly care whether they do or don't—all I suggest is that if you find people who have much in common—if you ask them enough questions and they give you enough answers in common—then maybe you'll find people who are liable to get along together, assuming that the questions that you ask have sufficient relevance to real life.

This is a strictly empirical approach toward matching people in terms of their likelihood of getting along with each other. It avoids at least the obvious pitfall of trying to employ such constructs as anxiety, empathy and dominance, which often are situational bound and which may be very transient.

DR. HAYTHORN: There has been a lot of interest in providing empirical tests of the old adage "birds of a feather flock together" and generally this holds, but it only accounts for a small amount of variance. There is another old adage that "opposites attract" and that also holds for particular variables, like dominance, for example. We found with our pairs in isolation that if you put two men together and both of them have a high need for dominance as measured by the Edwards Personality Test, for example, they had a great deal of trouble, at least in our studies at Bethesda. Similarity isn't always a good thing.

DR. RADLOFF: As an intellectual grandson of Kurt Lewin, I was of course, delighted to see you put his formula on the board, but also it gives me an opening for a comment I would like to make. The thing we are looking at is behavior, our criterion that we want to predict is behavioral in nature. It seems to me that everything we try to do in selection stems from this. As I said yesterday, it may look like putting the cart before the horse to specify in as much detail as feasible and

*Barratt, Ernest S. Factor analyses of some psychometric measures of impulsiveness and anxiety. *Psych. Reports*, 16, 1965, 547-548.

**Christie, Richard and Geis, Florence (1968) *Studies in Machiavellianism*. N. Y.: Academic Press.

necessary, the behavior we are trying to predict. The level of detail and specificity will vary from situation to situation. For example, for all Navy enlisted men, advancement in rate or retention as a criterion may be as much detail as you can manage. For submariners you may want to know more than that, with a little more detail about how they do their jobs, how they get along, and so on; for fighter pilots you may want to know their performance in the aircraft, and for bomber crews, how they work together. Dr. Gunderson and I have focused upon social and emotional behavior as well as performance in our work with the Antarctic and SEALAB data. These classes of behavior have to be specified in such a way that you talk specifically about what you are trying to predict. By working backwards and while still focusing on the group interaction dynamics in the context of the total situation one finally arrives at an assessment of the original selection variables. So, we must work backwards from the "B" in the formula $B = f(P, E)$ which is put first in the formula. I think this is where we have to start selection research, with the criteria. The criteria may be decided by commanding officers, but if the criteria are specified by the Command, they probably are going to have to be reified and spelled out in more detail by the psychologist. I think that it is of prime importance to specify what we are predicting.

DR. HAYTHORN: You may recall that I also put on the board a parallel third formula (S - O - R) that may be more familiar to some of you who aren't Kurt Lewin's grandchildren.

DR. WISKOFF: There are two points that come to mind. I was thinking very definitely in terms of S - O - R just before you put the formula on the board. One of the things that disturbs me somewhat is the concept of hypothetical construct, as Dr. Rimland, for example, presumably uses the term. Are we talking about motivation, leadership, or what? I never can quite get a feel for these terms. I think that sometimes psychology tends to reify the hypothetical construct, forgets that it really isn't observable, and that we are primarily interested in the stimulus

and response. We talk about leadership for example, and while this is a very important term in the military, I don't know what it really is. Is there something that you can pick up like an ash tray and say, "this is leadership?" We can define behavior and can define stimuli, call it leadership or call it another name, I don't care. I think we put too much stress on this type of concept. I will go along with what Dr. Radloff has said namely, that we should start with the behavior and go back to the stimulus, and whatever falls in between, if you want to call it something, well and good.

Mentioned also was the notion of personality variables, and we have given this a pretty tough go-around in the last couple of days at this Workshop. We've all acknowledged the fact that personality, background, character, whatever it is, is pretty difficult to measure. I want to make a comment related to the personality concept. This may have been said earlier, but it seems to me that one of the problems in talking about personality is the term personality itself. I don't think selection workers are really concerned whether it is personality they measure or what the term is. In a very empirical sense, it may be related to an aptitude approach that measures a cognitive realm or it may be biographic in nature. If it predicts something important, it is important. I think it is really the specific items in a personality scale that we should be concerned with, as to whether or not they have any predictive validity. The difficulty that we have with the concept of personality for individual prediction is surely multiplied in attempting to look at compatible or incompatible personalities, in terms of crew effectiveness. We're not even at the point where we know what personality means for a particular individual for prediction purposes. I think, in summary, I would put out a plea for empiricism in terms of S—R, or empiricism in terms of items that predict, and forget some of the terms like personality.

DR. HAYTHORN: I couldn't agree with you less. The fact is, I believe that you have to hypothesize some kind of intervening variable. If you are willing to predict what a man will do from one measure to what he will do

in an adjustment situation, you have to assume that he carries something along with him. If you can teach someone to respond in a particular way to a set of items and as a result, assign him an I. Q. of 140, but, in fact, he is an idiot or very close to it, you wouldn't accept that measure of intelligence as being predictive of behavior in another situation. You would, in fact, whether you said so or not, assume that there was something that the man was carrying around in his head that we call intelligence. Similarly, with electronics aptitude, clerical aptitude, or whatever else you measure that you want to predict, you have to assume that it is something that the man has within him, that is consistent over time, at least to some degree; so, I just can't buy the "nothing but" philosophy. In short, I think you have to be very careful about what your measures are, and you have to be very careful about making assumptions that go beyond your measures. I believe that when people say they are not talking about anything "but" the scores on the tests, they are just avoiding the issue.

DR. WISKOFF: I'm not disturbed about the notion of a person carrying around something in his head, something that the man may hold over time. This has to be the case, otherwise, we wouldn't be able to predict anything. If everything was very transitory and existed only for the moment, obviously we wouldn't be able to get any predictive validity. My only comment was reifying the particular term that we apply to this situation, leadership taken here as an example.

DR. HAYTHORN: Yes, people do get fuzzy with terminology, and I would agree in arguing that we need to be careful about that. But it seems to me that we have to assume (that is, if we are going to be in the selection business at all), that we are measuring something that has to do with the "O" or organism in the equation and we have two ways of doing it. We either measure stimuli to which "O" has been subjected—his learning experiences, his demographic variables, or things of that nature—or we measure responses—responses to specific tasks, responses to specific situations. But in both of those cases, whether we are at the stimulus or re-

sponse end in our measurement, we are trying to get at something that is inside "O" and I think we just have to be very careful about what it is we were talking about. It seems a reasonable assumption therefore that the response to a personality test is a measure of "O," if the test measures what it purports to measure.

DR. WISKOFF: My point is that by virtue of the processes we customarily refer to as perception we have gotten into some difficulties for philosophy. It would seem that the "S" and "E" never were removed from "O" and "P." How do we separate all of these factors?

DR. HAYTHORN: Professor Boring has said that every major advance in psychology has come from a better ability to measure the stimulus, and maybe that's right. But I don't really see how you could talk about selection unless you were trying to assess something that resides in, or is in some way referable to, the organism.

DR. HUNT: It seems to me that in the first two days of this Workshop we have been getting into more and more difficult problems. As we have proceeded, there seems to have been more and more reification of undefined concepts. On the first day, we had a number of references to specific variables, specific factors, and specific test items. In this session, I've heard of a number of tests, but with little or no mention of the item content. Doesn't anyone have particular specifications or behaviors or items that characterize some of these scales we have been talking about?

DR. WISKOFF: I don't care what they call them or what the nature of the items are if they predict some useful criterion. We are using experimentally a number of scales, for example, Kipnis' scale of persistence already mentioned in the Workshop session on personality tests. We are exploring the use of tests like this in predicting success in language training to see if perhaps some of these items would be predictive of success there. Also, I am told that tests of this kind are being used in the UDT training situation.

DR. RADLOFF: There is a statement somewhere, I think, in the Koch Handbook, something to the effect that, "personality theory

has become a rank, weedy garden or playground for persons of high fluency." I think this is the kind of thing we are objecting to.

DR. GUNDERSON: I have one example I'd like to offer. It may not clarify the issue, but I think it illustrates the problem. For part of our selection batteries for Deep Freeze, we had Len Gordon's Leadership Scale, which, as you know, presumably "taps" something in the area of interpersonal values. The content of these items involve dominant, directive kinds of behavior, traits indicating the degree to which a person likes to tell people what to do. This scale correlated something like — 0.40 with one of our peer and supervisor criterion in the Antarctic setting. The finding isn't difficult to explain. Most of the men of the station reported a somewhat egalitarian kind of atmosphere. Dominant kinds of leadership traits are inappropriate for the situation. Status leveling is very typical. So this type of measure actually nets us a negative predictive relationship with effective leadership behavior.

DR. NELSON: I would like to enter once more, the plea for better biographical information. The human being is a social organism and has experienced social situations throughout his life. We already, in our usual approach to personal history forms, include items that are related to social situations, but I think we could use some new ideas, and perhaps some ideas for better interpretation of those items which we are using at present.

I have three other comments. The first regards crew assembly. Once we surpass a two-man crew and we deal with three or more persons, we need a strategy for putting larger crews together. Dr. Haythorn mentioned earlier, the increase in complexity as the group size increases from two to three and more. There are in many of these groups positions and formalized roles to be fulfilled. Moreover, we can anticipate rather accurately which positions are, in fact, going to be functionally related, what proportion of the time, and in what capacity. This, of course, is one approach that Dr. Gunderson and I tried with the Antarctic stations and it seemed to work quite well. We didn't look at the assembly of the total crew, but began with the sub-sets

of the crew members who are formally expected to have the closest interaction. One question I'd like to perhaps have someone react to is the question pertaining to crew selection through self-selection versus assembly by other means. There is some literature, of course, which bears upon this matter.

Another problem with crew performance has to do with our attempts to predict performance on the job from individual test scores.

In the crew performance situation, we have contingencies of events with part of everyone's behavior somewhat dependent upon the behavior of other persons. It may be that we need to account more effectively for the behavior of other men in an integrated crew task situation to really evaluate the performance of any one individual. A good example was your story about the bombardier. If the pilot isn't over the target, not even the best bombardier will be able to make a hit.

DR. HAYTHORN: . . . Or even if he is over the target, but doesn't fly steady, a stable platform, he still won't be able to make a hit. There is the self-selection approach, in which men have been allowed under certain circumstances, to select their own crew mates. These groups were compared with random groups in similar conditions. When the conditions are right (and that's the tricky question), this technique works very well.

In teams of bricklayers, it has been reported to be the most effective single personnel change that a large construction firm in Chicago undertook—simply to let the two-man teams of bricklayers select their own working partners. Also, the approach was shown to be effective in small air crews in coordination scoring.

But the trouble is that in a typical selection situation where a lot of men are coming through and one has to form crews rapidly, it is difficult to give men enough exposure to each other to make reliable selections of each other. If you are dealing with a pair relationship, I think self-selection is a very promising procedure. In dealing with anything larger than that Roby has shown that there is very little transitivity. If "A" chooses "B,"

there is a good chance that "B" will also choose "A," but if "A" chooses both "B" and "C," there is no tendency at all for "B" to choose "C" in the kind of situations that Roby examined. So, that if you are dealing with larger than a two-man crew, the self-selection procedure gets to be less attractive. It may still be useful for combining teams of men.

With respect to the systems analysis orientation, I think what's being said in this Workshop Session is that a great deal of the variance is environmentally determined, and if you are predicting from scores of some kind obtained from the individual's performance, say under operational conditions, you have a lot of error variance as long as you aren't taking into account that portion of the variance that is environmentally determined. If one should try to do that, hopefully, he would be in a better prediction situation.

CAPT CHRISTY: With respect to the problems involved in the selection of two-man crews, does the notion that "likes" seek out "likes" hold? Possibly, this principle would account for the essentially monogamous nature of man.

DR. HAYTHORN: It could very well be. However, the mutual attraction notion breaks down often when generalizing beyond dyads. The fact that Joe likes Mary and George likes Mary is no guarantee that Joe is going to like George.

DR. WHERRY: I have a couple of comments with regard to crew selection. Some of the work that was done in Norfolk on the ASW crews in aircraft indicated that not all of the team members are equally important. Now this point was alluded to earlier when it was recognized that one cannot, meaningfully, talk about a five-man crew as being a single element or unit. It may be that two crucial members may be considered an element. The JEZEBEL operator for example, and the tactical officer in the ASW program, seen to be the most important crew members.

I also suggest that crew cohesiveness is a function of how long the members have worked together and may not, in fact, be a function of personality variables at all. One thing that the Norfolk studies indicated was that if you break a team apart (they were

not studying personality variables, but how long they had been working together as a team) you could not find another JEZEBEL operator and stick him in the team and expect as good performance. Similarly, you could not expect to keep the JEZEBEL operator and the other three members and pick up a new tactical officer and expect them to work as well.

The kind of data available might be illustrative of the kinds of habits characterizing teams that have worked together. Indeed, maybe there was no predilection to begin with to abort a mission, but having aborted one mission, maybe it becomes habitual, and having had to abort missions, maybe the motivation becomes lower.

DR. HAYTHORN: I think the answer to that is that this result has been replicated in the laboratory where that kind of thing has been carefully controlled. In this particular study, I don't believe that there is any systematic difference in how long the crew has been over there. I believe (I'm not sure of this) they were all crews that had returned from rest and recuperation, which meant that they were over there for half of the tour, whatever that was (sixty missions, I think) so they had been over there for thirty missions.

DR. RADLOFF: That crew stability data out of Norfolk is very impressive, but I think it does tend to mask some of the points that have been raised at this session. For example, if you look at motivation on the top, (this is the group norm), to find if it is high or low, cohesiveness just tells you the extent to which the group norm will be activated. Clearly, when cohesiveness is high, the group norm will be activated more.

Dr. Haythorn has indicated that these studies are approximate replications of laboratory studies done by Dr. Schachter, now of Columbia University. The basic concept is, the more cohesive the group, the more they will adhere to the group norm whether the group norm leads to high productivity or low productivity. The longer the people are together, the more they will adhere to the norms of their own group.

DR. NELSON: I think another point should be made, perhaps implied in the norm idea,

having to do with the cultural history of a group. There is a communications system operative in all of these groups which may not have a one to one correspondence with what is presented in training schools. A new man in the crew must develop, among other things, the informal language which is operative in that specific aircraft, let alone the squadron, or whatever unit of organization is appropriate.

DR. HAYTHORN: The Strategic Air Command talks about Crew Operating Procedures that are somewhat different from Standard Operating Procedures. A particular group of men will learn to do things a bit differently from the standard procedure. For example, a mission might be scheduled (this was a matter of squadron record), after which it could be aborted either because the crew got out in the plane and found the plane mechanically unsuitable, or they got in the air and discovered some failure that they thought was serious enough to turn them around and come back, or they came back because of bad weather, or they got over North Korea and they ran into opposition such that they diverted to a secondary target. Any one of those was classified as an "abort" in this data. It is important to note that the attitude measurement occurred during training before the men went overseas.

DR. KLAGSBRUN: Some of you were talking earlier about lineage, and the fathers from whom you have derived your present approaches. As a psychiatrist, I can't find my own father in the group of names the discussion groups have raised so far. All this calls to mind a major point involving the problem of professional communication. The background and information which I use and rely on in my work, in spite of the fact that I had a psychological background before a psychiatric one, creates a gulf between us. I refer to a different set of data, a different type of material than the material that you people are dealing with. And yet, in spite of that, I feel that we're talking about the same problem, namely, group dynamics.

For example, there are data related to the subject matter now under discussion available from the Washington School of Psychi-

atry and the Tavistock Clinic in England. Moreover, the problem of group interaction has been closely examined in some of the two-week confinement studies conducted in this country. People from different backgrounds have been put together and have gone through a number of manipulations including dividing a very large group into subgroups without an agenda, or having the subgroups create an agenda with inter-group exercises. These are fascinating things to do if you are studying groups or sub-groups in any major organization such as found in the Navy. It would seem that these approaches provide an excellent way for people to learn about themselves as well as a means to train for group situations in which they may have to function at a later date. A submarine is, in a sense, an ideal group process laboratory. However, should members of our discipline and contemporary social psychologists separately study the group dynamics of submariners, the communication barrier that I mentioned earlier would again quite likely appear. This may, in part, be the result of prejudice on both our parts—but, more likely, results from the fact that we don't have ready access to each other's data.

DR. HAYTHORN: This kind of data and the research situation sound interesting. I would wonder, however, about generalizing to real life situations.

DR. KLAGSBRUN: Let me cite one example. The setting for this type of work emerged from certain of the industrial problems in England. The main function of the Tavistock group was to find ways in which industry, and this included those that were nationalized, could set up communication systems so that a person who was on the Board of Directors of one industry could understand and communicate effectively with another person on that same board without coming to blows. Incidentally, I suspect that when industry became more nationalized, these problems became more acute. In short, the need for this approach actually came out of a realistic problem rather than from an academic research problem.

LT GREEN: An intriguing statement has been made a number of times during this

Workshop Session. Namely, that we should select the man for the team. This seems to be an extension of a related concept, that of selecting the man for the job or the task, though the team selection idea probably involves much more complex processes.

Dr. Bob Wherry and I have been discussing a somewhat different problem in semi-privacy during the last day of the Workshop. At the Naval Missile Center at Point Mugu, his responsibility has to do with "Naval Technical Evaluation," which involves, I believe, testing the weapons systems with the crew on board. It seems that rather than just testing the system itself in isolation, it is a prime opportunity for the social psychologist to do the one kind of research which I think should be done.

From the Workshop discussions so far, it has become apparent to me that psychology has become too departmentalized. Psychologists have become sub-specialists, and sub-specialists just don't talk to one another. I think a lot of the research each of us is doing could benefit from the thoughts and ideas of the others. This Workshop seems to be providing a means for these information exchanges and demonstrates, I think, the mutual benefit of the experience. We should forget that I'm a social psychologist, and someone else is a physiological psychologist, or whatever. Some of the concepts, techniques and methodologies that I feel comfortable with and quite frequently use in my research might be of benefit to someone else interested in a slightly different area but is not aware of the concepts.

DR. HAYTHORN: I think that the different situations that we find ourselves in probably account for a great deal of the differences in points of views we have. I can have more faith because I'm not involved in having to come up with any selection procedure for next month's Navy, or next year's Navy. We're looking ten to fifteen years ahead at Bethesda, and so we can keep asking, "What will be the characteristics of men, and teams of men?" that far off and we can, at least for a while longer, cling to the faith that maybe we'll develop an ability to compose crews rationally, and this may then make a

significant improvement in our ability to predict individual and group behavior. If you have to worry about what to do about prediction next month, or next year, you can't fool with that kind of luxury. You have to go back to what are the measures and what are the correlations that we have right now. I think that this kind of meeting may serve the purpose of getting those of us who have a different mission in touch with those who are dealing with the real world as it now is. This ought to benefit both of us.

DR. SCHAEFER: The problem of inter-person compatibility seems very much to the point in an age of exploration into the gaseous outer space and the liquid inner space beneath the sea. While our work involving circadian rhythms emphasizes, for the most part, the periodicity of physiological functions, nonetheless, we have, in the process of conducting several confinement studies, made some observations which shed some light on this matter.* One key concept here would seem plausible, namely, that inter-person compatibility may be some complex function of the degree to which the cyclic processes (both physiological and psychological) of the persons involved are "in-phase," or for that matter "out-of-phase," over a period of time. Though of necessity, our human sample size for our studies has been small, nonetheless, there may be some suggestions for meaningful research into these matters contained in some of the Submarine Medical Center literature.

LCDR GALLAGHER: I am wondering after a day and a half of this workshop, what information I am going to carry away with me. Have we solved any important problems, or have we simply raised them? I have two points to raise, both obviously on different levels of discourse. First, I think our group processes work should attempt to identify the factors that account for the adjustment strategy that works in one group but not in another, the group observations to include data pertaining to the nature of the situation

*Schaefer, K. E., Clegg, B. R., Carey, C. R., Dougherty, J. H. and Weybrew, B. B. Effect of isolation in a constant environment on periodicity and physiological functions and performance. USN Submarine Medical Center Rpt. No. 488, 1967.

in which the group is functioning. Secondly, when I look at some of the research being conducted in some of our laboratories, I am forced to ask myself some specific questions. For example, "What weapon system is going to involve two individuals in long term confinement?" I feel that research in Navy labs should have relevance for the Navy, and the two-man isolation work taken as an example seems to be of doubtful operational significance at this time.

DR. HAYTHORN: One comment that should be made in response to the first point is that studying variables that are relevant to a particular situation does not mean that you have to simulate that situation exactly. As a matter of fact, if it did mean that, we'd be out of business because it's impossible to simulate a situation exactly.

LCDR GALLAGHER: My question is, "Are we dealing with the real world when we study two-man teams in long term confinement?"

DR. HAYTHORN: No, we are not dealing with the real world. We're specifically not dealing with the real world. We're dealing with what we hope is a representation of variables that will be important in some real world situations.

LCDR GALLAGHER: But I can't conceive of the weapon system that's in the offing that will involve two men in long duration confinement.

DR. HAYTHORN: While they don't stay out for extremely long periods, there are two-man aircraft and two-man submarines. One can conceive of the weapon system where there will be isolation from society, relatively small crews in situations in which men will interact.

DR. WHERRY: As an example, one might look to the JEZEBEL operator and tactical officer as a relatively effective dyad. Possibly total crew composition is not as important as dyadic composition. There is a unique aspect to the example found in the fact that periods of duty as a dyad seldom exceed twelve hours. Possibly most any pair can get along for this long.

DR. HAYTHORN: I agree with this, but

look at what you're saying. I think you are saying that variables interact and moreover, unless you have exactly replicated the real world situation, you can't generalize at all from the laboratory situation. If that's the case, we can never generalize from any situation because no two situations are exactly alike.

One has to argue now that the degree to which you can generalize from one setting to another is a function of the similarity in situations which is something we can't describe very adequately because we don't have any adequate way of describing situations. But, what we're assuming in the work we're doing at Bethesda is that the systems like the NR-1, systems like a SEALAB Habitation, systems like SINBAD, all have in common small groups cut off from the larger society for relatively long periods of time. Now that's the cluster of problems with which we have been charged and given the mission to study.

Now, one way of doing that would be to simulate a SINBAD capsule. Nobody can tell us what it looks like, nobody, because it's just a gleam in a bunch of engineers' eyes. What we know about it is that contact with the larger society will be limited, crew size will be small, and there won't be much to do other than leg work. We can look at those kinds of variables. We know that we'll have Navy sailors in them, and we know that Navy sailors have certain characteristics, and we would like to study the effects of some of these characteristics. Now, I would be the last to argue that you can generalize from a laboratory situation directly to a field situation. But, obviously you've got to be able to abstract to some degree if you're going to do any research at all.

DR. SCHAEFER: To ask that the laboratory situation be essentially the same as a field situation may not really be relevant. In the laboratory situation, we are simply seeking to identify the basic characteristics and processes involved in the phenomena we are interested in. I will give you an example: We studied cyclic processes in several two-man teams in the laboratory. We have shown, among other things, dissociation of

a number of functions resulting from the removal or flattening of time-givers. Having found this in the laboratory, we then go to the submarine to look for the cyclic phenomena. To conduct applied research, you need basic facts to be applied and these come from laboratory experimentation.

DR. HAYTHORN: It seems to me that we come to the question of preferences for various methodological approaches in every conference of this kind. The argument usually involves the differences between field research, as opposed to exact simulation, as opposed to laboratory study. It seems to me that these lie along the continuum of increasing abstraction, and generally, what Dr. Schaefer said is accepted - i.e., the more abstract situations allow you to get at more basic aspects of the problem. You then assume, partly on faith, but backed up by a lot of experience, that the laboratory findings will be of use in the future.

My own feeling is that the Navy needs a wide spectrum of approaches and that we need to interact. We need to get experience from the field to grind into the design of our laboratory studies. We need an opportunity to take our laboratory findings and test them in a field setting. But to say that any one of these approaches is the only way to go, I think, is wrong. As Dr. Schaefer says, in field research studies you have little control, you have limited access to observational data, you have a great wealth of interacting variables, the relative significance of which is unknown. Obviously, it's very hard to get reliable information in that kind of a setting. In the laboratory setting, on the other hand, you can get more reliable information but you lose some of the ability to generalize to the real world situation. The selection researcher needs both basic and applied information.

DR. NELSON: Well, I think what is called for is more systematic communication between those who are focused on the field situations and those in the laboratory. Those studying in field settings can provide certain life situation cues to groups in the laboratory which they have not been able to build into their experimental situations, but which might be interesting for them to contemp-

late. On the other hand, laboratory groups are able to provide some insights which we might evaluate and use in field situations. We do need better concept definitions for effective interchange between field and laboratory study groups. To interchange between the Antarctic studies and those in the NMRI Lab, for example, a common focus upon confinement as a concept might be better than upon isolation.

DR. HAYTHORN: Well it's a different concept. We're really looking at three rather distinct aspects of the situation: the stimulus reduction, which seems to be generally a boring, comparatively monotonous situation; the social isolation, being cut off from a larger society; and finally, confinement, which is quite a different source of stress. The three are not necessarily correlated.

DR. CHAMBERS: I might add, that within any given group of personnel there are many types of working crews. These crews are composed differently, depending on the type of work activity or task performance requirements. One of the problems in selection is that we have some crews whose members perform one way on certain types of work tasks, and another way in other types of work tasks. The role of individuals and the structures of the crew change according to the activities in which the group or crew is engaged. Selection of crew members is very difficult for each of these changing roles.

For example, about fifteen years ago we conducted a study involving several sociometric variables in a situation in which routine work activities, assigned scientific activities, and recreation activities were the three major categories of work to be performed. We were studying crew composition as a function of performance requirements. We found that crew composition, i.e., crew structure, varied as a function of the types of work activities being performed. This turned out to be an extremely important observation. The people who were the center of one particular crew cluster were rarely, if ever, the center of the other clusters. They were sometimes the isolates, and the crew was organized differently. Here we have another dimension of the problem area, viz.,

the kinds of activities that are required by the crew.

DR. PLAG: I don't think that there's anything I can add with regard to the P-E interaction. However, I would like to raise an issue about the variable interactions. I've looked at thousands of interactions and I can't find anything that's uniquely predictive. I'd like to ask workers in similar areas if they have done the same thing, and if they have found many interrelationships of useful predictive value.

DR. RIMLAND: We have done several studies in which we find the sort of thing described earlier in this session in regard to Kipnis' work i.e., when you divide the group on one variable, you find greater predictability on other variables. For example, we found in nineteen different Class A schools, a zero correlation between a man's Navy Knowledge Test (NKT) score and his likelihood of re-enlisting four years later. However, we found that by dividing the NKT into thirds (upper, lower, middle) and running validities separately for each third against an attitude scale toward re-enlistment, we could use the NKT to improve prediction of retention in eighteen of the nineteen groups. It was those who knew most about the Navy who were better able to state whether they were likely to re-enlist in the Navy four years later. But even so, the correlation jumped from an average .07 to an average of .19, which isn't very much, although the increase was very constant. We have found this sort of thing in several instances. There are some non-linear interacting situations in which prediction is improved by taking advantage of dichotomizing or trichotomizing a predictor but, we haven't found anything very powerful so far.

DR. HAYTHORN: Dr. Kipnis' findings were replicated in the sense of controlling for a variety of Navy rates. He had samples of subjects who were electronic "techs," hospital corpsmen, machinist's mates, and so on. He found the same general results in all of these samples. If you call this procedure replication, then he has found these interactions predictive. The results were not rep-

licated in the combat setting, but were in the laboratory setting.

DR. HESTER: I've been recently assigned to the task of developing a weapon system effectiveness model. For security reasons, I can't talk about the model specifically but a few comments about the nature of the problem in general might be of interest. In the first place, when you are planning a new system you have to ascertain what is possible, including possible criteria. In some instances when you put this system into effect you may find that your plans for criteria do not work out. As a result a new or revised set of criteria must be adopted. In a word, the system evaluation problem may be quite different from the system design and development program. So, my thinking would appear to be closer to the statements made earlier in this session by Dr. Nelson and Dr. Wherry. My first step in systems evaluation would be to try to take out the environmentally determined variables. If there is reason to believe that the role of some individual is crucial to system performance, we might attempt to study the effects of the individual on the system, that is ascertain how much of the outcome can be predicted from his unique contribution. This approach treats the group interaction as a residual. In a sense, don't deal with it until you have to. I have been accused of leaving the people out of the situation I'm trying to study, but I don't think that's the case. I think by doing this, you thrust the real problems on yourself and a lot of the overwhelming complexity will disappear because now when you determine that the system does not perform effectively, the only explanation is that the crew is not fully compatible. And when you've got that, then you have a situation in which you can observe the interaction and attach some meaning to it. But it seems to me that if you begin with the group process, at least in analyzing an operating system, you end up with a situation that's so overwhelmingly complicated that it is difficult, if not impossible, to comprehend.

DR. KLAGSBRUN: From a clinical point of view, I've always been impressed by the fact that if you use that approach, you end up

studying a different being or phenomenon. In psychoanalytic research, the same problem exists. Many analysts who study groups probably as a result of their theoretical background do not take into account the fact that the group is a different entity than the people who make it up.

DR. HESTER: That's true, except in a military situation you deliberately impose control because you know you are likely to have a varied group, you know that you can't predict the people you're going to get, and yet you know you must somehow predict the outcome, so you deliberately set up a control situation in which you subordinate most of the group. This is not just to be nasty or to be dictatorial. It becomes an unnatural group, but in that process much of the interaction affects are brought under control. In effect, problems are eliminated because you have a series of people responding to a leader in terms of their individual tasks. Now this is not entirely true, but you move in that direction—that's the whole point. It's possible that in an informal group in which the predominant interaction is allowed to take place freely, leadership emerges naturally. Under certain circumstances it might be the most powerful, most effective way but you'd never trust it in a military situation.

DR. HAYTHORN: I think if you're analyzing a specific situation, you're quite right. You ought to start with the criterion measures and work backwards until you've accounted for at least some of the variance. Our problem is a bit different.

DR. HESTER: I know it is. I'm beginning to see the point. It's when you're developing the system that you have to at least establish the plausibility that such a system can in fact operate within certain limits before you build the hardware. If you can show that it is plausible that groups of this composition can actually function, you've in a sense, accomplished your mission.

DR. HAYTHORN: I think we can say something about some of the variance. What proportion of the total variance is accountable in a real life situation remains to be seen.

DR. HESTER: True, you explain as much as you can, but you need not feel defeated

because you can't account for it all. You can show that the group is workable. Your account of some of the variance in detail is, in a sense, a bonus. To repeat, showing that a given group composition is workable is an accomplishment. This is the big hurdle, possibly the biggest one.

DR. HAYTHORN: I would like to say more than that—that the social-emotional adjustment and the performance effectiveness of such groups can be modified to some degree by directing attention to group composition.

DR. HESTER: In principle, there would appear to be no question. Yet, showing the effect of group composition in the performance of a complex operational system can be most difficult. For one thing, people with acute emotional problems, those most easily distressed by adverse influences with the group are often dropped early in training. People with extreme characteristics simply won't do. This is one problem. At the same time, if the system is complex, cognitive questions may account for a major part of group to group difference in performance. Under these conditions, isolating group composition effects, however real, is seldom very useful in a practical sense—but anything you can contribute is helpful. I don't think it is necessary to account for all the variance.

MR. MOLYNEAUX: If this approach is "pushed" far enough, one can end up engineering the group. I would say this is a useful direction in which to go in this area.

DR. HAYTHORN: I hope we're beyond that. Hopefully, we are in the stage now when we are talking about engineering groups on some rational basis.

DR. MOONAN: I would like to say two things. First of all, I was very interested in your classification of the group processes called "forming, storming, norming and performing." I noticed that these events have been occurring during this conference, particularly during this session. A very informative "forming" discussion has developed and I think there is plenty of evidence of "storming" going on. How much "norming" is taking place between sub-sets of this group is very difficult to assess. I would like

to address my comments to the "performing" process.

My interest is in mathematical and operational techniques that one would utilize if he desires the information that you are striving to determine. I am embarrassed to say that I don't know of many methodologies, but I will tell you of two that come to mind.

The first is a technique devised by Professor Paul Dwyer* some years ago which provided a method of assembling aircraft crews. Is this utilized in any of your efforts? The second methodology I refer to was written by mathematical economists, Professor David Gale and Dr. L. S. Shapley.** The title is something like "The Marriage Problem and the College Placement Problem." I am sure this reference would be of interest to many present at this Workshop Session. I have heard the phrase, "forming groups of two in a confined environment," and that is an appropriate characterization for the marriage problem! It does contain a very simple and clearly written rationale for the "crew assignment" problem where "crews" consist of pairs of individuals for marriage and where they consist of college students for academic placement. I might also say, from what I remember of it, that the procedure is not transitive in the sense that if you assign colleges to students or husbands to wives, the assignment is not the same as assigning students to colleges or wives to husbands. You can take whatever policy suits you, based on other criteria. I am interested in the possibilities of the newer mathematical techniques, particularly linear programming, dynamic programming, and decision theory. The discussions during this session of the workshop have motivated me to look into the problems being confronted by selection workers generally. My major point is that you have to make assignments under some kind of restriction. You don't have all of the men with all of the desired properties, so you have to make trade-offs in which you can at least identify and hopefully specify a func-

tion or functions. The technique should be fairly straight forward, although I can't point to a specific methodology at this time.

DR. GUNDERSON: It is apparent that those of us concerned with group composition need all the help we can get. I am happy to hear that Dr. Moonan, a mathematician, is interested in this problem and since he is located at the BUPERS field unit in San Diego near by our BUMED Neuropsychiatric Unit, I can benefit from his counsel on statistical technology.

I would like to comment on the points that have been raised about the relevance of our concern about group composition. Recently, I met with the Assistant Chief of Psychiatry and the Task Force Medical Officer at the Bureau of Medicine and Surgery to discuss the selection of Deep Freeze groups for the coming year. In that context, we discussed the possibilities of experimentally composing groups. For example, how would we select from a pool of candidates the ideal group for year long duty at the South Pole station? This is an insoluble problem with our present knowledge and "state of the art." Also at that time, I visited Dr. Haythorn's group at Bethesda. They were similarly asked for suggestions as to how one would go about this selection task. Of course, they couldn't tell us how to do it; however the opportunity to "cross-fertilize" our ideas was quite beneficial for our group interested in "field" approaches as well as for the Bethesda group with their laboratory approaches. We agree in general with Dr. Haythorn that gains in our capabilities to predict group compatibility and achievement, even with our crudest available tools, will soon be possible.

As I have already mentioned, we have been trying to measure both group compatibility and group effectiveness. We are presently looking at the composition of these groups in terms of all the screening data. One interesting fact in relation to the Deep Freeze selection problem is this: for most selection situations, we are selecting for a rather stable organization or institution, where norms and work roles are firmly established. You can obtain samples, determine who is successful or unsuccessful and cross-validate

*Dwyer, Paul S. The problem of optimum group assembly, AFTPTRC-TN-56-18, Lackland Air Force Base, Texas, 1956.

**Gale, David and Shapley, L. S., College admissions and the stability of marriage, Rand Corp., 1961.

the results generally. The situation at the Antarctic stations, and probably in certain other groups, is one of a completely new and unique group, which forms at a given point in time, exists for a given period of time, and then finally dissolves. In this situation, you have differences in group values, or attitudes and group composition—all of which are quite important for the group structure. We have been, and still are, intrigued with some of the observed differences in the value systems of the groups and differences in the compatibility and achievement in relation to the variations in group composition.

CAPT CHRISTY: I just have one question addressed to the discussion leader for this session, or to anyone who may have some information on the matter. The question is simply, “do volunteer groups differ from non-volunteer groups? If so, in what ways?”

DR. HAYTHORN: I'll tell you one thing about them. In the stimulus deprivation work that we've done, we've had subjects sent down from Great Lakes without volunteering. When they get here, we test them and then ask for volunteers. Those who volunteer are used as subjects; those who don't are sent on their way. One significant difference between the volunteers and non-volunteers is that the volunteers are higher

on a measure that Tom Myers calls “thrill seeking.” This is a preference for adventurous kinds of activity, through new experiences and the like. What he further finds is that that score is negatively correlated with endurance. This comes back to the comment that Dr. Gunderson made this morning about motivation—asking a man how much he wants to be in the situation may not be the best indicator of how well he will do in it. Here we have a case where the same measure positively predicts volunteering for the situation, and negatively predicts enduring in it. By enduring I mean lasting through seven days without asking for early release. As to other differences, I'm not sure. We have a high volunteer rate, so we're not eliminating many subjects at the volunteer level. We have about 85% volunteers. Obviously it is difficult to say much about the remaining 15%.

If there are no additional questions or comments, I'd like to terminate this session with the observation that these discussions have demonstrated rather severe and only partially solved methodological problems related to the measurement of both “P” and “E” in the hypothetical equation with which the session began.

SESSION VI

PERSONNEL ALLOCATION PROCESSES

William J. Moonan, Ph.D.

U. S. Naval Personnel Research Activity

San Diego, California

This presentation is concerned with the general aspects of personnel allocation processes. As participants in this conference, you are intimately involved with the management and design of such processes. The group has indicated its interests by the papers presented and discussions that have so far taken place. The Workshop program indicates that there is yet more on this subject to follow.

The contents of my discussion will indicate that personnel allocation processes are a particular case of general decision theory. Special sub-cases such as selection and classification will be identified and defined. Emphasis will be on the information required to be collected and produced as well as the costs of carrying through personnel allocation programs. An illustration of an allocation model which uses cost information will be provided.

On a related matter, two recent computer programs developed at the U. S. Naval Personnel Research Activity will be discussed. These psychometric programs are used for selecting items for inclusion in a test score or a battery of test scores. These programs have indicated that test lengths can usually be increased by item selections determined by these programs. Since psychometric data are frequently used in personnel allocation programs, these computer programs are particularly important.

We frequently are obliged, in our operational work, to allot personnel to one or more activities. Obviously, there are many ways to do this even though there may exist certain constraining conditions. The procedure of allocating resources, in such a way as to maximize (or tend to maximize) the effectiveness of the assignees when they are assigned under any constraining conditions, is called an **allocation process**. Such processes are special cases of more general **statistical**

decision-making processes. In the field of operations research analysis many different types of allocation processes have been identified. Efficient computing methods for these processes have also been developed. Many of these use the techniques of linear, quadratic, dynamic and heuristic programming.

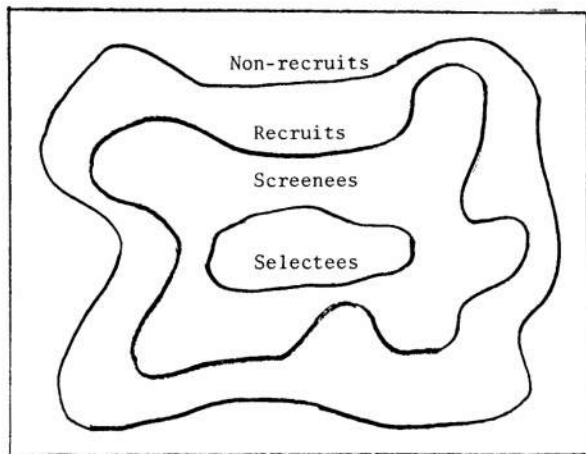
In the personnel field, special names for different types of personnel allocation processes have been given and these have been used freely in this conference. In order to provide a common basis for understanding these special terms I will give operational definitions of them and provide illustrations. This effort may provoke comments from you in the discussion period. If so, we shall all profit from any further clarifications that evolve.

The personnel research sub-cases of general personnel allocation processes are variously referred to as selection, screening, classification, testing, categorization, assignment, allocation, recruiting, counseling, graduation, etc. Let us consider operational definitions of some of these terms. The definitions are: **Recruiting**. The process of inducing personnel to apply for admission to some program. Those personnel who are influenced to apply for admission are called **recruits**. We shall also use the term "recruits" to refer to those personnel who are to be allocated by any of the special sub-cases of a personnel allocation process.

Selection. The process of identification of those recruits, who meet certain selection requirements is called **selection**. In practice, selection requirements are not always fixed initially but are conditional on situational circumstances. In this case several stages of selection, called **screening**, may be involved. During these screening stages, selection requirements may be relaxed or tightened.

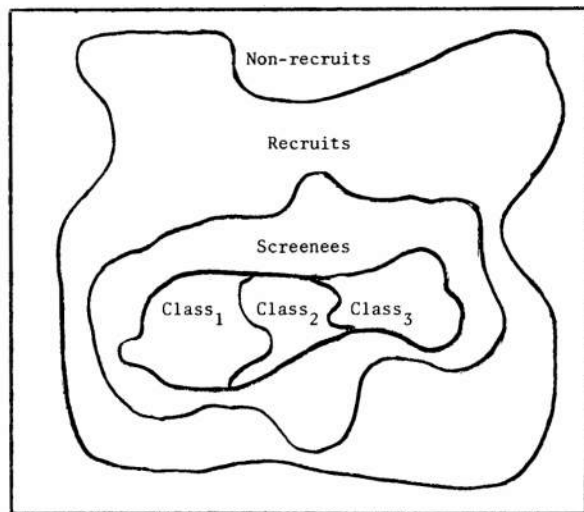
The terminal stage of screening is usually referred to as the selection stage.

We can illustrate the aforementioned terms in the following diagram.



Note that the ultimate objective of selection is to decide which individuals in the population are to be selected for a single type of activity. This screening process divides the recruit population into two sub-sets, namely "selected = S " and "not-selected = \bar{S} ." More specifically we might have the sub-sets labeled as trainable or non-trainable, retain or not-retain, and hire or fire. These dichotomies are characteristic of the selection process since selection is intended to allocate recruits for a single category, namely the selected category. The complimentary sub-set is often not defined.

Classification is defined as the process of allocating recruits into two or more categories. For example, Naval Recruits are classified at the Recruit Training Center into one of about 90 class A schools or the Fleet. Classification, in any particular problem, may be preceded by recruiting and several stages of screening. The following diagram illustrates the process.



The distinction between selection and classification probably needs to be emphasized. A selection implies that a recruit is allocated to one selected group. A classification implies that a recruit is allocated to one of several groups.

The term "classification" has meaning in another type of allocation process. Consider the existence of a sample of recruits each of whom is measured on several variables. We then can consider two mathematical "spaces." These are the **variable space** and the **sample space**. We might wish to partition the variable space into two or more sub-spaces. Multivariate statistical techniques known as cluster analysis and factor analysis are appropriate strategies to employ for this purpose. On the other hand we may wish to partition the sample space. The techniques of object-cluster analysis, pattern recognition and Q-factor analysis are procedures used for this purpose. If the criterion for affecting the sample space partition comes from the information inherent in the variable space, then the process employed to effect the sample space partition

is also called "classification." The process of partitioning the sample space using criteria external to the data is called **discrimination**. Once the sample space classes have been established, the statistical techniques known as discriminant function analysis or multiple discriminant function analysis are employed as allocating processes for recruits.

Testing is the process of evaluating recruits or trainees in order to obtain information regarding their knowledges, abilities, attitudes, physical conditions, etc. In certain special types of testing a decision process is built into the testing procedure. This occurs in sequential testing where the test space is partitioned into sub-spaces called ACCEPT, REJECT or CONTINUE. Usually an independent selection or classification process follows after each testing is completed.

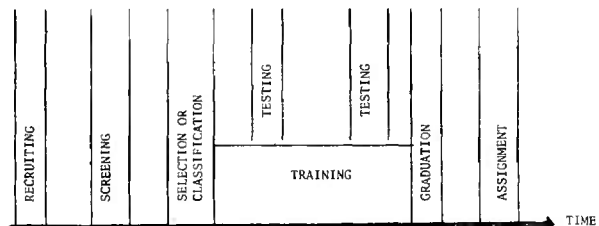
Graduation is the process of allocating trainees into the special sub-class called, for example, "successfully completed training or orientation." The graduation state is one near the end of an allocation process.

Assignment may be considered the terminal state of a personnel allocation process. In assignment we allocate graduates (or non-graduates) into specific billet assignments or duties. Upon occasion no duties are involved, but graduates are merely made eligible for other programs and allocation processes.

The following diagram is presented to represent the allocation stages associated with a typical Navy personnel allocation problem. The diagram is not intended to represent any particular process or program but is intended merely to portray the sequential stages involved in the typical process of the type we are associated.

The allocation process requires that information of various types be provided or acquired at all stages. Furthermore, at all stages, certain costs must be borne in order to carry out the program. Initially the management of a program must make many decisions. Not the least of these concerns the characteristics of the population from which recruitment procedures will be attempted. Various instructions, pamphlets and other descriptive information must be prepared and disseminated. The cost of this can be

considerable. The process of recruiting is costly especially if it involves the **interviewing** process. Also information, specific to the recruits, is acquired and recorded. Screening processes are ordinarily not expensive since most or all the information about the recruits has been processed and only various selection strategies are to be experimentally applied to these data. Dr. Gunderson has discussed this problem as it applies to the Antarctic program. Selecting or classifying individual recruits, if the number of recruits is not too large, is also not too expensive. In special cases, or if large numbers of recruits are involved, computer programs and algorithmic developments may raise costs.



We have considered **training** as a component in a personnel allocation process. It is important to note that the greatest costs are typically involved with the training sub-process. It is for this reason that prior allocation sub-processes must be carried out extremely well. If not, we face great loss associated with the costs of training personnel who will not be effective in those activities to which assignments are made or, in fact, cannot be assigned. It is, therefore, very critical that effective recruiting and screening take place. For this reason the policy of graduating only a certain quota or proportion of trainees, independently of their training attainments, seems unwise from a cost point of view. A policy of graduating a certain proportion of trainees can result in two types of error. Errors can result by graduating inferior trainees or not graduating satisfactory trainees. If the number of either of these errors is large, the overall allocation policy needs to be critically examined. The trainers should bear the problems of motivation and training,

but only the minimum of problems associated with selection errors.

During the training stage, information is gathered, at some expense, which relates to the degree to which training has succeeded with the trainees. This frequently takes the form of psychometric or other types of testing. It is important for this testing to be accurate, since if it is otherwise, poorly trained personnel can escape through the graduation sieve and be assigned to duties they cannot effectively perform. Although activities to which assignments are made have their own safeguards for this problem, occasionally very expensive or drastic events can result due to the presence of poorly trained personnel in the field.

As we have seen, the costs involved in allocation are numerous and varied. Sometimes they are hidden and they are almost always difficult to evaluate. As research workers we would all like to forget about them and pretend the costing problem did not exist. We cannot do this. The administrative managers of allocation programs cannot do this either. Costs must be accounted for in our allocation models. Information requirements are also numerous and varied. We have problems about what to collect and how to use it. One benefit of this meeting results from hearing what others are doing about solving these problems. Modern statistical decision theory is our most useful mathematical tool for incorporating the cost factors.

One primary objective of the allocation process is to fulfill a manpower requirement. This must be done by providing a certain number of competently trained men at the assignment stage. Furthermore, as Mr. Molyneux will justly insist, we must do this at minimal cost. A model for doing this will now be discussed. The model chosen was originally devised by Kao and Rowan (1). I will omit many details which can be found in the reference. This model illustrates certain useful points and is not too complicated. Incidentally it was one of the first attempts to bring the cost concept into the personnel allocation process. The model over-simplifies the allocation process we have discussed but represents some of the issues in setting up

selection decision models. It will not illustrate the process of allocating individual recruits.

The problem we will consider requires us to specify a cutting score of a selection test. This cutting score must be that which is appropriate to insure that the probability is $(1 - \alpha)$ that the number of trainees graduated will be sufficient to meet the total assignment quota denoted by n . We will assume that we have a single selection test, X , and a single criterion score Y which we dichotomize into two classes; graduated, not-graduated. The situation is diagramed below.

		Y		
		GRAD	NOT GRAD	
X	PASS	$\pi(1,1)$	$\pi(1,2)$	$(\pi 1, \bullet)$
	FAIL	$\pi(2,1)$	$\pi(2,2)$	$(\pi 2, \bullet)$
		$\pi(\bullet, 1)$	$\pi(\bullet, 2)$	1

The $\pi(i,j)$'s represent the probability that a recruit selected at random from the set of recruits will fall into cell (i,j) . It is also assumed that the joint distribution of X and Y is known. Perhaps we may assume a bivariate normal with specified parameters. The variables X and Y must be assumed to be correlated, otherwise X would not be useful for predicting Y . We need not worry about the case of perfect prediction! Given a cutting score on the criterion variable, Y , we can then determine the marginal probabilities $\pi(\bullet, 1)$ and $\pi(\bullet, 2)$ from the distribution function information.

The marginal probabilities $\pi(1, \bullet)$ and $\pi(2, \bullet)$ represent the probabilities of a random recruit passing or failing the selection test, X . These probabilities can be chosen by establishing a cutting score X_0 . We would not care to choose X_0 arbitrarily but rather with some purpose in mind. Our purpose shall, of course, involve costs.

The costs we shall consider include the costs of recruiting, testing and the loss incurred by selecting a recruit who will not graduate. We can aggregate these costs into a cost function of the form

$$C = C(N) + [N\pi(1,2)]L$$

where,

N is the recruit set size, $C(N)$ is the cost of recruiting and testing N recruits, $\pi(1,2)$ is the probability of not graduating a recruit who passes the selection test X , L is the cost (loss) accrued for each such recruit, and $N\pi(1,2)L$ is the total cost of these selection errors.

Mathematically our problem is to minimize the value of the cost function subject to the condition that the probability is $(1-\alpha)$ that the number of graduates is greater than or equal to the number required for assignment. Symbolically we state the condition as

$$\text{PROB}[N\pi(1,1) = N(1,1) \geq n] = 1 - \alpha$$

We require an estimate of the number of personnel to be trained to be assured these requirements will be met. The degree to which we wish to assure this depends upon our protection level $1-\alpha$. If this is set very high, e.g. $1-\alpha = .999$, we can be well protected against a shortage of graduates. We pay for this through the cost function term $N\pi(1,2)L$, however. If $1-\alpha$ is set low we may not be able to fulfill requirements for men without lowering graduation standards.

Omitting many details and comments provided in [1], the analysis leads to a specification of the optimum value of $\pi(1,1)$ and the corresponding cutting score of the selection test. Of course the remaining values $\pi(1,2)$, $\pi(2,1)$ and $\pi(2,2)$ are of interest and easily obtained once $\pi(1,1)$ is given. All of the analysis is informative and should be executed with various values of the parameters $N, 1-\alpha$, $\pi(\bullet 1)$, $\pi(\bullet 2)$, and distribution function parameters of X and Y .

Generalized mathematical models for personnel allocation processes are relatively rare. In order to work in that field, a researcher must be competent in mathematics, statistics, operations research, psychometrics, as well as have a very good command of statistical decision theory. As a consequence of these requirements, plus a few practical ones, it is not too surprising that extensive modeling procedures have not been devised. I would appreciate having my attention called to recent developments in this area. Perhaps I

can conclude my discussion with some references on computers and mathematical models. Before coming to this conference I visited Harvard University where I learned about a new book on the subject of personnel classification (6). Although I have not seen the book, the reputation of its authors would imply that it contains valuable material for us. Dr. Robert S. Ledley (2) has written a large volume on computers, mathematics, statistics and other subjects that will surely be useful to you. One interesting application concerns the use of multi-stage decision theory for the special allocation process called **diagnosis**. A final reference, among many that are appropriate, is a recent book by N. T. J. Bailey (3) which develops mathematical models for biological and medical processes. In this regard, I recently learned that Dr. Richard Bellman, one of the most creative mathematical geniuses this country has produced, has recently left the Rand Corporation and joined the staff of the University of Southern California. He is expected to work in the mathematical-medical field, and one can expect dramatic developments as a result. With this I must leave this part of the discussion and devote the remaining time to the presentation of some recent computer programming developments which have occurred at the U.S. Naval Personnel Research Activity.

Reference (4) describes a computer program, called SEQUIN, which has great potential value for all stages of allocation processes which utilize psychometric tests. As applied research workers we are all aware that good criterion variables are very hard to come by and that relationships between these variables and selection or classification tests is generally poor. Consequently any improvements in validity of the predictor instruments would be welcome. Attempts to increase validity can generally be classified into one of three strategies. One of these involves abandoning old tests and seeking new ones with higher validity. Another involves combining old and new tests. Such techniques as multiple linear regression analysis are useful for this purpose.

SEQUIN provides an example of a third strategy. This strategy uses the old tests but

develops new scores from them. This procedure has the virtue of reducing testing development costs. It also has two other virtues which may surprise you.

Numerous examples have shown that the "test scores" proposed by SEQUIN were very much shorter in terms of the total number of items used to formulate scores. Also the validities for these shorter tests have always been larger than the validities for the longer tests. In some cases the difference between the validities was very remarkable. The consequence of this is that testing time can be reduced or more tests of different types can be given in the same time. Moreover, overall validity will almost surely increase. Costs will be reduced because more selection errors can be avoided.

A brief description of the program will now be given. Using an input-form and item response information from personnel, a sum, sum of square and cross-product matrix of the item responses and criterion is created in the computer. Using this information the program then computes the item validities. The item which has the greatest validity is selected as a major component of the test score. Having done that, the procedure finds another item, which combined with the first, is associated with the two item test score of maximum validity. This sequential process is repeated until the available item set is exhausted. If we have I items, there are $I(I+1)/2$ iterations undertaken. A printout is provided at each of I selection stages giving the number of the item selected, test mean, test standard deviation, reliability and validity information. The graph of the validity coefficient as the number of items increases appears below.

The maximum of the validity curve has usually occurred between 5 and 25 items. The drop off of validity of course depends upon the sample, the criterion and the item set. In some examples the drop has been dramatic (e.g. $\max=.80$, $\min=.25$). The Navy Basic Tests have generally shown relatively small drop-offs, but quick build ups, for a variety of criteria.

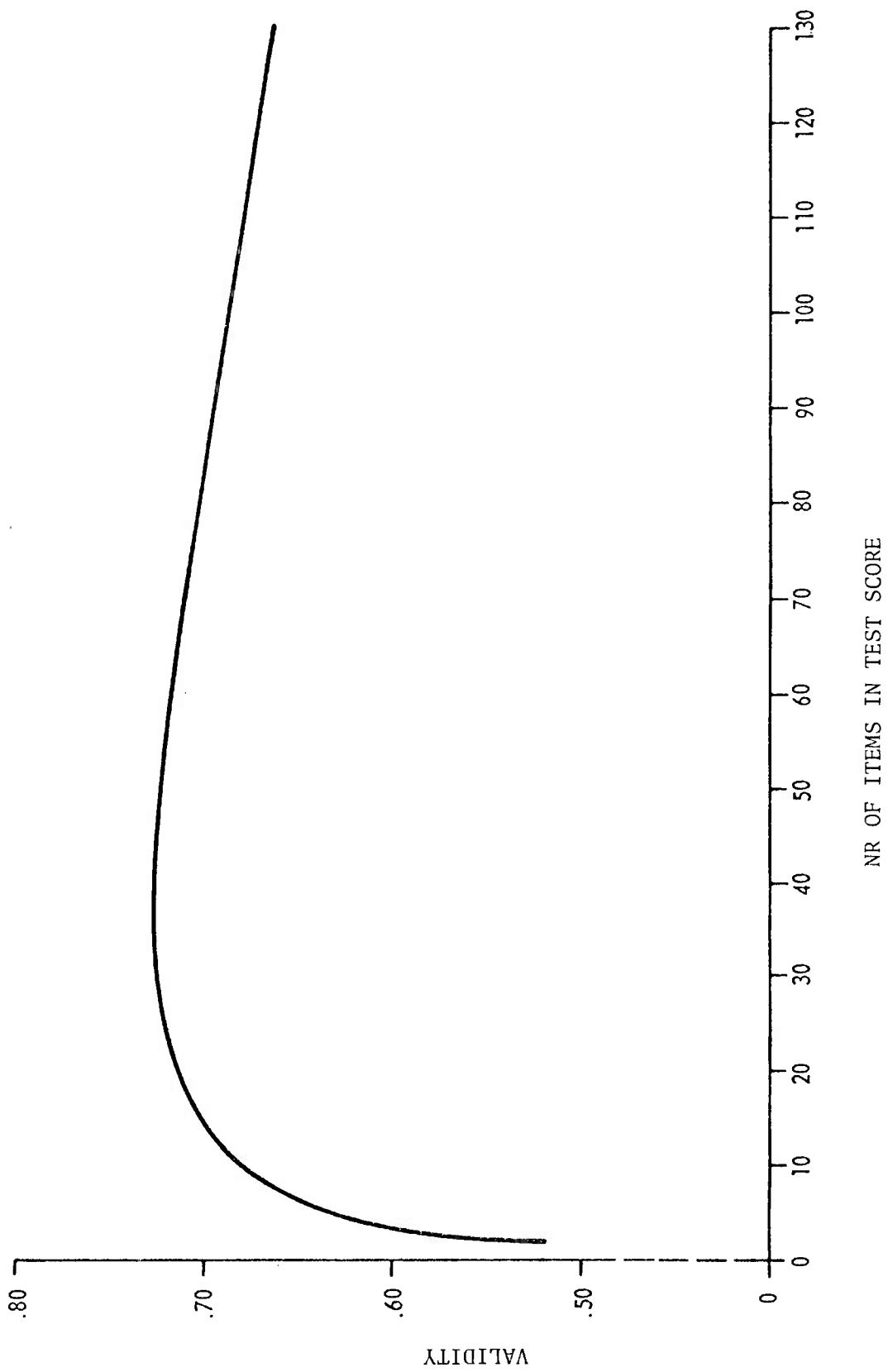
The second and last program I shall mention is called SEQUIN II. This is a general-

ization of SEQUIN. The generalization consists in extending the number of tests involved in the analysis from 1 up to a maximum of 9. This program makes it possible to increase test battery validity. The program sequentially selects items and accumulates these into the scores for each test for which the item is associated. A multiple regression equation is constructed which maximally predicts the criterion from knowledge of the test scores. The item selected at each selection stage is that one associated with the greatest increase (or least decrease) in the multiple correlation coefficient.

The program is now operating and a report, reference (5), will be issued soon. Very similar results were obtained for one case where SEQUIN II output was compared with that of SEQUIN. In that case all available items from 3 tests were combined in a single test for the SEQUIN run. This result indicates that inter-test item relationships were similar to intra-test item relationships. If this were not true, SEQUIN II should lead to better prediction resulting from use of the same number of items. Present programming restricts I to 180 for SEQUIN or SEQUIN II. Programming for $I < 600$ is under way. NPRA believes that, through the use of these programs, validity coefficients can be significantly improved with or without using extra test information. This result should improve the personnel allocation programs we develop.

F. REFERENCES

1. Kao, R. C., Rowan, T. C. *A Model for Personnel Recruiting and Selection*, System Development Corporation, SP-15, 1958.
2. Ledley, R. S., *Use of Computers in Biology and Medicine*, McGraw Hill Book Co., New York, 1965.
3. Bailey, N. T. J., *The Mathematical Approach to Biology and Medicine*, John Wiley and Sons, New York, 1967.
4. Moonan, W. J., Pooch, U. W., *SEQUIN: A Computerized Item Selection Procedure*, U. S. Naval Personnel Research Activity, SRM 67-8, 1966.
5. Moonan, W. J., Balaban, J. G., Geyser, M. J., *SEQUIN II: A Computerized Item Selection and Regression Analysis Procedure*, U. S. Naval Personnel Research Activity, (To Be Published), 1967.
6. Rulon, P. J., Tiedman, D. V., Tatsuoka, M. M., Langmuir, C. R., *Multivariate Statistics for Personnel Classification*, John Wiley, 1967.



FREE INTERCHANGE FOLLOWING DR. MOONAN'S OVERVIEW

DR. MOONAN: It seems appropriate to start the discussion period by having Dr. Gunder-son elaborate on the selection system used in the Deep Freeze Project and Dr. Radloff to proceed similiary for the SEALAB Project. I would like to know, in particular, if they feel that the general personnel allocation model that I discussed is applicable to these projects.

DR. GUNDERSON: First let me give you some numerical estimates regarding the size of the population "pool" from which we can recruit. Let's confine our consideration to one group that must be selected for Antarctic service, namely, hospital corpsmen. As you know, the men must first volunteer for Antarctic duty. Incidentally, and it is not known what this means, there are fewer men volunteering since the Vietnam situation arose. We have to select five hospital corpsmen to winter-over at stations in the Antarctic. We start out with a population of approximately 19,000 hospital corpsmen from which to "pick" these men.

Each year the Navy circulates a notice to all ships and stations requesting people interested in this duty to apply. It describes the duty, specifies the rates required, and instructs them to apply through their local commands. We may have 100 hospital corpsmen who initially apply to their local commands. At that level, a medical examination is completed before forwarding the man's application which is accompanied by a rather detailed medical history. Let us assume further that twenty-five people are either denied forwarding of his application or given a negative endorsement by his Commanding Officer, or they have some medical problem which is specifically disqualifying. That leaves seventy-five potential applicants. The Bureau of Naval Personnel then orders a number of these men—those selected as eligible candidates—to special screening centers. As you know, the Bureau of Naval Personnel has certain standards for rejecting candidates, such as not having been paid a travel allowance within the past year, not having serious indebtedness and not having sufficient job

experience. So we may end up with twenty-five men being sent to the screening centers. Here special examinations are given, including intensive medical and psychological testing. Out of the twenty-five at the screening center, let us assume that twenty are considered fully qualified. I'm exaggerating the number disqualified at the center slightly; however, the task for the medical officer or administrator is to select the best possible corpsmen from among the twenty qualified for this special assignment. There are eight different Navy enlisted ratings or occupational groupings to be considered in a similar manner. I wonder how Dr. Moonan's model would apply to these situations?

DR. MOONAN: A model of the type discussed does appear to me to apply to problems of this nature; however, I prefer to go into a detailed analysis of a specific selection problem at this meeting, since the analysis for any individual case is different and time consuming. There are many complexities and details to discuss. To do so now would not use the conferees' time effectively.

Dr. Radloff, would you discuss the selection procedure used for SEALAB?

DR. RADLOFF: Unfortunately, I don't have detailed knowledge of SEALAB selection. All I know is the general approach to the selection of the aquanauts used as subjects in this study. These procedures would seem generally to characterize programs of this kind in their early developmental stages. You have, to use a sociological term, a charismatic leader or two, and they often are highly experienced in the particular field they are working in. They know a large number of people in that field; they attract people to them, and they select from within the group of men with the greatest potential for the mission at hand. Here selection is based upon length and depth of experience of the candidates. Dr. Moonan's model does emphasize several important points, one having to do with the fact that often selection and classification occur simultaneously. For example, among the divers who are eligible for the program, men were chosen specifically on the basis of certain capabilities to perform specific tasks related to the SEALAB missions. They were

classified and assigned to the experiment in exactly the numbers needed. During training one SEALAB II aquanaut was eliminated from the program for one reason or another. He was immediately replaced by another man with very extensive experience, comparable to that of the man he replaced. Quite likely, if other men had been dropped out, they would have been replaced in essentially the same way. The point of relevance for selection is that ordinarily there is not an excess of men chosen for the study prior to training. They ended with the same number they started with in the case of SEALAB II.

DR. MOONAN: The wisdom of using that strategy is conditional upon the costs associated with having dropouts in the program.

DR. RADLOFF: I agree, particularly if the term "cost" includes the psychological cost to the dropouts, that is the stigma attached to a man being eliminated from the program. The public image factor resulting from the publicity certainly affects the aquanaut's motivation. Many of them have strong attachments to the program. It would be very difficult to choose thirty-five men and eliminate five of them at the end of the training period for example.

I want to point out that exactly the same thing happened in the first space mission. They choose seven astronauts and six went up; the only reason Slayton didn't, was because of a heart murmur.

DR. MOONAN: One needs to realize that it is a traumatic experience for most men who become dropouts in the various programs. You have a useful way of characterizing the effect as "psychological cost." I would hope that we could evaluate this cost and make effective use of this information in the allocation process.

DR. RADLOFF: The Naval Academy has an attrition picture somewhat like this. Flight training at Pensacola is yet another example of the same thing.

DR. MOONAN: Dr. Weybrew, would you discuss the selection problems encountered in your work?

DR. WEYBREW: One characteristic I like about Dr. Moonan's model is the one just

mentioned, namely, that it takes into account the cost of attrition, costs not only in terms of dollars, but also in terms of the quantity and quality of human resources being removed and retained by the selection procedures. As I indicated the first day of the Workshop, the submarine service enjoys a relatively low failure rate for both officers and enlisted men at the training level. It varies from time to time, but is of the order of ten to twenty percent for enlisted and from three to seven percent for officers. No serious problems arise until we get to the point we call the career decision point, namely, when an officer's obligated service is over or when an enlisted man's enlistment term expires. At this point, the first re-enlistment rate varies from twenty percent for some electronics ratings for example, to sixty percent or more for certain clerical rates. On the other hand from fifty to sixty-five percent of the submarine officers extend beyond obligated tours of duty.

DR. HUNT: About twenty-five years ago we used a model something like this. It would seem to be a fair question to ask what the model really contributes. The inputs to the model are not new; however, it does involve mathematical formulae that will enable one to program a computer to do a number of things in connection with selection and classification matters.

If you look at basic training realistically, however, you'll find that graduation from basic training is simply a matter of fitting the original criteria that were set up. Although admittedly circular, the only people who don't graduate represent errors in the original selection procedure. In other words, any recruit who truly fits the recruiting standards will complete his basic training.

What one can do is to take the original selection discharge rate at the training level (the screening rate) and over-complement to compensate for it. Instead of 100% complement crews coming, if your discharge rate is three to four percent, one raises the recruit quota to 103 to 104 percent of complement, knowing that somebody will "pick up" the extra three to four per cent and they

will be "bedded down," not in training station space, but in a psychiatric unit, a medical ward, or in the provost's office.

Then you get to what, for me at least, is a very interesting question. What is basic training? If the only recruits that fail to complete it are the ones who should have been caught in the original selection procedure, just what is the purpose of basic training? This question is never faced; somehow it is lost in the model. In subsequent trade schools, training becomes quite real and people fail to complete it. What is the meaning of basic training to a recruit if the only ones who don't graduate are the ones who don't meet the original criteria? It isn't a true training function. This discussion obviously leads to the conclusion that basic training is really a further selection procedure and does not in a sense have training as its major function.

DR. WISKOFF: It is like any other selection procedure. We know that our selection will not be exact and there is going to be attrition during basic training, whatever the category or reason.

DR. HUNT: I'd like to throw a figure in here, if I may. The cost in the early days of World War II for men with an inaptitude discharge (instead of being picked up at the induction center) was about \$650. This is what it cost to send him to the training station, give him \$20 or \$50 in his pocket, a new suit of civilian clothes, and send him home.

CAPT. CHRISTY: One thing that concerns me is the fact that some of these criteria aren't being applied in ways that seem most useful. A man may have a high GCT so he is placed in a school requiring (for example) a top secret clearance. Finally, he is assigned a billet at an isolated station to continue the example. He turns out to be sexually or otherwise immature and he breaks down because somebody didn't take a long enough look at his personality when he was recruited. That is why we need to get sub-criteria into the "picture" somehow. I think some of the things that John Plag has found are quite useful. On the other hand, they ought to be used as a "flag" to call attention

to a particular man to be observed more closely. For example, if a man's father died when he was eighteen, it's a lot different than if he died when he was seven. If he has a good substitute father, he may be all right.

Another thing often forgotten in this context is the cost of courts marital, brigs, mass confinement and "BCD's." Quite often, the doctors involved don't even see these people who were separated from the Navy because of a character disorder for example. Oftentimes, moreover a non-medical Veterans Administration person has given him a fifty percent rating and the taxpayers are stuck with paying the pension.

One final comment seems indicated. Suppose that the man has been in a top secret school, like nuclear submarine school for example. If he is dropped after assimilating all of this top secret material, we will be in for trouble, or could be, particularly if the man is unstable.

DR. WISKOFF: I think you've got to look at the realities of the situation. Some of the work that John Plag has done argues that a more effective job of screening could and should be done prior to boot camp.

I think we're all aware of the current "push" from the Department of Defense to take in greater numbers of lower mental level personnel. This works completely against what we, with our scientific (or not so scientific) methods and tools know we can do. That is, a considerably better job of screening these people out could be accomplished, but we're not allowed to. I think we have to work within the realities of the situation. One of these realities is that we will not be able to screen these people out very effectively before they have gone through some sort of training which will necessarily result in a high attrition rate at that level. We just have to live with this as long as we're following that DOD policy. There are allocation systems where the objective can be stated as a problem of minimizing the number of poor quality men.

DR. MOONAN: I should reemphasize that my illustrations of allocation models are not

completely general. The most general formulation is very complex. For one thing, it involves consideration of the personnel system external to the particular area of application. The important point to be emphasized is that the effects of any suballocation system are felt more or less in the whole personnel system. Therefore, the nature of the whole personnel system needs to be considered in setting up a specific model.

LCDR WHERRY: In most instances, I doubt that we will ever be able to prove whether we made the right choice or not. I submit that there are hundreds of ways to make those decisions and we may never know whether a given choice was the best; you may only know if it was wrong.

Some time ago a DOD group came to Pensacola to examine the feasibility of the notion that there would be hundreds of thousands of dollars saved in the gas that these men (who later dropped out) would have burned up during the training program. But the whole process is really based on the fact that we cannot collect enough valid information during the initial selection period to make hard and fast determinations as to who should stay in training. I think we all ought to be aware of the increases in the amounts of information that we do get on men once this initial decision is made, namely, that a man looks like a good prospect and so on. You should not stop your selection there, but go into what we may call secondary selection; that is where the real pay-off is.

You want to identify the floaters early in the game and get them out before you use up too much training cost. I think that there are also some bad elements to this. Let's assume one can devise a seemingly very effective scheme as to how you are going to drop these so-called poor risks out of the program at an early date. But unless you can get the Command also to buy the procedure, too many applicants will have been taken into the program and your training cost will go up.

There is a third point that I want to mention and that has to do with the quality of the "product" making up the training

input. We don't know how DOD or SECNAV really wants to play the game. Is a good jet pilot worth twice as much as a good multi-engine pilot, or a good helicopter pilot? One cannot minimize costs or maximize the quality of the total output from the classification standpoint unless this information is spelled out. In short, I am saying that the training command should not be making these decisions and yet they are making them.

MR. MOLYNEAUX: Actually, they shouldn't be. Your decision is made for you as a result of the requirements established by CNO.

LCDR WHERRY: Your decision only says that you will have so many people in this category and so many in that one, it does not give specific qualitative requirements in all cases.

MR. MOLYNEAUX: That is true, but the attrition rate is going to be something that is being decided by both the selection procedures in effect and the training requirements.

DR. RIMLAND: It is not enough simply to set the quotas. You actually have to set a quantitative evaluation on how much these people are worth in terms of a trade-off formulation of some kind.

DR. MOONAN: Except that these people are different and you have identified this difference.

DR. RIMLAND: Not in terms of value. As Cronback and Glaser have pointed out, underlying each personnel decision are certain, generally unrecognized, assumptions concerning the values of various outcomes. The Navy wouldn't pay ten million dollars to acquire an additional helicopter pilot, yet it certainly would pay ten thousand dollars. Between these values, somewhere, is the value, in dollars, of the pilot. When you decide where to assign a man who has a forty per cent chance of passing helicopter training and a fifty per cent chance of passing jet training, you are making implicit judgments of these values, whether you realize it or not. It's not enough to know you have a quota of so many jet pilots and so many helicopter pilots. Unless you make

your value assumptions explicit, there is no mathematically sound way of proceeding.

MR. MOLYNEAUX: This approach seems to assume a "huge" population — something approaching an infinite "N." But we don't have an infinite "N." Instead, we have a limited "N." When one has a limited number of men to select from, one can only try to meet the quotas. One attempts to distribute your available talent so that the probabilities will be maximized toward meeting quotas as they have been calculated or at least estimated. Then you set up cost effectiveness comparison charts, and make your selection based on where the lines cross.

DR. RIMLAND: This implies that you know some actual quantitative values. You have to know what kind of decision errors are most costly to you in order to assign individual men in such a way as to minimize that cost.

MR. MOLYNEAUX: Let me restate my position. If you have a limited number of people to assign, you are going to assign them so as to achieve optimal use of the people available to meet your quotas. To do this, you must know the probabilities of success of the people in the various specialties and assignments. If we don't know the probabilities of success in various assignments, our procedures will not work. We need this quantitative information.

LCDR WHERRY: We have equations which allow us to predict the relative quality of the training grades if we put a given man say, into jet training or if we put him in multi-engine training. We can, indeed, maximize the effectiveness of the whole assignment system on the basis of giving out the overall best training grade. Now this would appear to be one method of maximizing both cost and effectiveness. In short, in Naval Aviation at least, the average training grade is the best available criteria on which to base these decisions.

On the other hand, what if a good jet pilot is really worth three times as much as a good multi-engine pilot? Should I not weight this jet pilot with a good grade three times as heavy in my system than the multi-engine

pilot with the good grade? What I'm saying is that these decisions have to be made at the CNO level; they cannot be made at the training level.

MR. MOLYNEAUX: You have introduced an additional factor no doubt of value to the Navy. But this type of problem is rarely raised by the selection people as a problem. So, unless CNO is aware of this type of problem, how can decisions regarding the value weightings of various specialties be assigned?

DR. WISKOFF: It is implicit in the classification model we've been talking about to know who is making these decisions. The trouble is that CNO and BUPERS have abrogated their responsibility by letting the classification and testing people make these decisions. We have to state what we want these decisions based upon; that is, we have to state what the function is that we want to be maximized before this can be put into a model.

LCDR WHERRY: I've been in the selection business for a number of years and it really didn't become obvious to me until a couple of years ago that indeed, this information had to be in the model somewhere.

DR. WISKOFF: We have been working for some years in San Diego toward automating classification procedures. As you know, many textbooks on selection present situations in which, for example, an employment manager may have job openings for say half a dozen machinists. With a large number of applicants, it is a relatively simple matter to select from the top down. In general, when the selection ratio is low (i.e. many more qualified men than jobs) most selection techniques have some degree of effectiveness. However, the classification problem as it relates to Navy school assignments is more complicated. Instead of having just one job opening, we have quotas for something like sixty schools and instead of having five or ten applicants for a few jobs, you have 100,000 recruits a year. At each training center, there are about 1,000 men a week being classified into these sixty jobs. This poses an extremely difficult classification

problem. What techniques, mathematical or otherwise, are most effective in handling the allocation of these numbers of men to these billets?

In the past, this classification task was done manually in a very clumsy, ineffective manner. The job was simply given to a classifier who, with knowledge of the required quotas, would simply take cards representing men and sort them into pockets, representing jobs. His task was to satisfy quotas with people who meet the cutting scores. Since there are millions and millions of possible combinations to meet these quotas, it isn't likely that the classifier would reach anything close to the optimal solution to the problem. Recently, we have been trying to do classification tasks of this kind in a much more sophisticated way.

Dr. Moonan and Dr. Wolfe (Dr. Moonan's colleague at the BUPERS Field Activity in San Diego) have developed a program designed to handle this type of classification problem. This program has actually been put into practice. Since October 1965, the classification at San Diego has been done by means of a computer rather than by hand. At the present time, Decision Systems, Inc., is under contract to develop a similar program for Great Lakes. Once both of these programs are operational, data from the training centers will be combined. This centralized classification task will be accomplished at Rockville, Maryland. By combining inputs and quotas into a pool, we'll get a better chance at maximizing the effectiveness of the classification system.

In one of the first trial "runs" of the new classification system in use at San Diego, we took as a sample, a group of 905 men (one week's input) with school quotas for 545. The manual classifier tried to match these men into jobs and he came up with 485 who actually met the cutting scores. Waivers had to be given for the remaining sixty men. In other words, they didn't quite meet the cutting scores and he had to lower the cutting scores for these men in order to assign them to schools. We took the same group of men and ran them on Program OPERATE, one of the computer programs

in the classification system. It turned out that we were able to fill all 545 slots. There were no waivers required and the average level of ability, as measured by the test scores, was slightly higher than that accomplished by the human classifier. This is an example of a resounding success for the method.

A problem that turned up later was that a great deal of the classification was really being done at the recruiting level where people were being promised what school they would go into before they joined the Navy. Therefore, a great deal of the potential of this model was being wiped out by conflicting policies. We are trying now to identify the policies that have been obstructing the proper utilization of computerized classification and trying to talk management into eliminating them one at a time. The recruiting people, of course, say that they need the capability of promising applicants certain kinds of training, otherwise they won't join the Navy.

DR. HUNT: It seems obvious that the computer is being accepted in our culture, as has the notion of data banks and the like. It seems a reasonable prediction that Navy recruiters will soon be telling the recruits that they will receive assignments in which they will be "best off." We need facilities dedicated to the purpose of making these assignments most effective.

LT. GREEN: I've been through this recruit training procedure myself including selection and testing. The last couple of days I've gotten the impression that many of us tend to see ourselves as omniscient people in our isolated ivory towers. We seem to forget that we work with human beings. We talk of a man as an experimental rat, an experimental unit. He is not. We discuss the tests and these test batteries over and over. But, how often do we look at them in the recruit training station? A Chief, who couldn't care less, administered the tests I took. He threw us in a room with noise outside, bad lighting, bad ventilation at 6 a.m. In affect, what we did for the next few hours with these test papers determined what would happen to us for the next four

years. No matter what we as social scientists do, we should never lose sight of the fact that we test and observe people, and from decisions made from these test data we control, in a sense, what happens to them for four years in the Navy. For that matter, often perhaps indirectly, we make decisions affecting the persons involved the rest of their lives.

If it is any comfort, there are a few statisticians and methodologists who see the problem and are trying to do something about the matter. There is a very serious attempt being made to improve the whole system. The computer, incidentally, is making this possible. The processes involved in recruitment, classification and career retention should be improved as a result.

DR. MOONAN: The model that is to be used next year for classification of recruits from San Diego and Great Lakes is a more sophisticated and better model than the models currently in use at San Diego. However, we still have a considerable way to go. Improve-

ments for the so-called "front end" of the model are particularly required. I believe we need to experiment with an adaptive classification procedure. We need one that makes a series of trial classifications. Since the recruits do not have access to the qualifications and motivations of their peers nor to the quotas, information to be used in the assignment process often is not adequately determined. Outputs from trial classifications can be used for feedback to the recruits where subsequent vocational counseling and adaptation can occur. Hopefully, the process would converge to the satisfaction of all concerned. We need also, a better vocational guidance system and training for the recruits in problems relating to military vocational decisions and implications of these decisions for their present and future lives. The use of time-shared computer systems, computerized counseling and guidance programs can be very useful if supplemented by appropriate professional psychological personnel. Thank you for your valuable and informative contributions to this meeting.

SESSION VII

SELECTION-RELATED RESEARCH AREAS AND DIVISION OF LABOR AMONG THE NAVY ACTIVITIES

SESSION OVERVIEW

SELECTION-RELATED RESEARCH AREAS AND DIVISION OF LABOR AMONG THE NAVY ACTIVITIES

LCDR Paul Nelson, MSC USN
USN Bureau of Medicine and Surgery

The concept of professional integrity was mentioned by Dr. Hunt in his keynote address and has been implicitly and explicitly relevant throughout this Workshop. I would like to use that concept as a referent for the remarks I wish to make this morning.

Whether our attire be military or civilian, and whether our work lies in the laboratory or the bureau, I look upon our common role as one of behavioral scientists jointly serving the Navy, a single, large, historically-unique and important professional organization. As my post-doctoral tutor Dr. Walter Wilkins likes to say, we work for the graduates of an engineering college on the Severn River and we must always keep this in mind. Our responsibilities are perhaps best characterized in nature by the term "advisor-consultant." Conventionally we are presented with broad goals and are asked within certain limitations to offer advice as to how to most effectively accomplish such goals simultaneously protecting the fitness and integrity of man and enhancing the effectiveness of naval systems.

With regard to the type of counsel we are capable of providing, I have been impressed throughout this Workshop by the broad range of professional backgrounds and interests represented by the participants. Perhaps as a result of this diversity one might come away from the present Workshop with the impression that many of the specific technical problems related to selection were not discussed in any great depth. At the very least however the present Workshop did provide a "broad spectrum" approach to selection-related problems with the common variance of the participants being their con-

cern with man as a human being, emotional, spiritual, physical and however else we may wish to characterize him. Manpower management in the broadest sense is a concept encompassing everything we have discussed in this Workshop.

In the past few days we have tried to focus specifically on the concept of personnel selection. We've, at times, mentioned training and have acted as if our collective consciences were saying, "Let's not get into training; this is a Selection Workshop." We have touched upon human engineering problems having to do with the manipulation of the material environment and with ecological problems of man at work in the Navy. We subsequently left that topic because of its seemingly foreign nature, re selection. Our problem is that we can scarcely talk about such matters as selection without becoming involved with additional concepts, one such concept being training. In all, I quite agree with the point just made by Dr. Moonan that whatever we call it, our problem is essentially one of providing advice pertaining to decisions, continuous decisions made over time with regard to the human population of interest. Those decision-making processes are called by different names depending upon their function in the personnel system. Recruitment, selection, training, classification and billet assignment are some of the "labels" often applied to these procedures. The common denominator for these concepts is often missed with a very narrow point of view regarding personnel selection being the resultant. It is certainly our professional responsibility to alert others to the similarities and differences among the many event categories which characterize our work and our role vis a vis naval service personnel.

Another problem which concerns me is the small amount of effort directed toward devising new conceptual schemes and research strategies in the field of personnel selection. We do have available some theoretical work,

but largely done by academicians. However some of the participants of this Workshop have made some important theoretical contributions in this field. New strides, for example, are being made in classification and assignment models. But we need more concentrated effort in the modeling area, and some of our approaches will have to be based upon modifications of our philosophies regarding manpower management. Perhaps, for example, we should be less concerned about collecting so much detailed information regarding the input to the Navy, that is, the classical procurement problem—and think more in terms of maximizing manpower utilization processes within the limits of the manpower resources available. Given any population input into the Naval Service, an input which may be based as much on political as on scientific merit, we should ask ourselves in what ways we can most effectively utilize that “pool” of resources so as to optimize the individual’s adjustment and performance as they contribute to the effectiveness of the Naval Service as a whole. I am sure others have thought in these terms; however, often we must be reminded of the responsibility to continuously pursue these ends.

One idea occurred to me during one of the previous Workshop Sessions when Dr. Wiskoff was describing the AFQT distribution existing at the present time. As you may recall, he described the distribution as being essentially bimodal with disproportionately fewer Mental Group III’s being represented. I first abreacted to this whole notion until it occurred to me that perhaps the Navy doesn’t need many more Mental Group III’s than the number it is presently receiving. If the AFQT is usefully valid, there must be information regarding the differences in performance capability between Mental Group II’s and Mental Group III’s and between Mental Group III’s and Mental Group IV’s. If, having delineated these differences, we could then match aptitude with job requirements, we might find that in a relatively technical organization such as ours, the optimal distribution on the AFQT is, in fact, bimodal in nature as it presently appears to

be. There are some types of jobs which require relatively low levels of cognitive functioning, but at the same time, require reasonably steady, routine performance, jobs which perhaps fit the performance potential of certain of the Mental Group IV’s. The remainder of Navy jobs may very well skip to the Mental Group II level of required cognitive functioning. I don’t know. I merely raise the point for thought. Again, we should pursue such matters.

Allow me now to comment on the topic I was scheduled to discuss this morning, namely, the division of labor. Our problem, I think is not so much one of division of labor, but one of division of man. To use anatomy charts as a model we have created a BUPERS impression of the Navy man, a BUMED section of the man, an ONR section of the man, and so on. We do have, I think, a necessity for some division of labor to pursue the various aspects of man’s functional characteristics. But again, we should be aware of the arbitrary categorizations we have created and communicate among ourselves and with others so as to emphasize the wholeness of man.

The major concern of the Bureau of Naval Personnel and the Bureau of Medicine and Surgery has been with man in his totality. The Office of Naval Research, of course, is concerned with a man in the context of all functions also. Many other Commands or Bureaus are concerned with man primarily as he becomes a part of specific hardware systems.

The first order responsibility for more effective communication therefore lies among the Office of Naval Research, the Bureau of Personnel and the Bureau of Medicine and Surgery, particularly the latter two. There should be no argument with regard to BUMED’s responsibility for the health of Naval personnel. What we do as behavioral scientists under the rubric of health is of necessity, focused strongly upon psychophysiological and social-emotional (i.e. psychiatrically related) characteristics of behavior. I think it is also fairly clear that BUPERS is concerned in the main with

problems of vocational interests, aptitudes, job and skill requirements.

But, insofar as the two broad classes of performance criteria on which all naval service personnel are evaluated, namely, conduct and proficiency — BUMED and BUPERS have a joint interest. To the extent that conduct denotes characterological adjustment, BUMED has been interested, to the extent that proficiency denotes fulfillment of job requirements, BUPERS has been interested. But both criteria must be examined in reference one to the other. It doesn't do much good to have a highly proficient man if he is not going to be reliable — if we can't depend upon him. Thus we must coordinate our efforts and knowledge.

I think through media such as this Workshop, we have initiated more effective communication, at least on an informal level. We must have interchange among bureaus and between bureaus and laboratories, and among the laboratories themselves. Quite possibly, one of the major benefits of this Workshop has been that it has provided an opportunity to become acquainted with both the professional staff and the research programs of other naval activities. This Workshop has brought together in one room professional researchers from different bureaus and activities, providing ample occasion for the interchange of ideas. We need more opportunities for these interchanges.

Finally, let me mention the quite time-consuming paper work designed to outline our research efforts, the forms known as DD-1498's. These forms are filled out initially by people such as yourselves in the laboratories. They are sent to the bureaus where they may or may not be edited in one way or another. If care is not taken at the bureau level, serious errors in classification of various research programs can, and occasionally are, made. This happens largely because of a lack of clarity in preparing these documents. For your information, the DD-1498's you prepare are used for library reference, for financial support of our research, and for planning of future developments within the Navy, often by persons with no background in our fields of specialization.

May I enter a plea therefore that we all put more effort into the preparation of the DD-1498's originating from the various activities, so that we, at the bureau level in turn can be more effective in presenting the programs to other interested people.

I would now like to conclude with the comment again that as professional psychologists, we have an obligation to one another and to the Navy to regard man as a total organism even though we may individually be responsible for only certain aspects of his behavior at any given time. In every manner of fulfilling that objective, we must communicate more effectively.

I have personally benefited from this Workshop. I would like to thank the Submarine Medical Center, specifically Dr. Gell and Dr. Weybrew for their efforts on behalf of the Workshop, and to Dr. Hunt for lending the dignity of his office and experience to our sessions. May I finally express appreciation to each of you who crossed bureau lines and even disciplines to participate. I hope we may have more meetings of this type, perhaps with different combinations of participants, perhaps some smaller group workshops with a little more intensive focus on specific problems, and perhaps some workshops which are organized more along "symposium lines." I personally will do my best from the Bureau of Medicine and Surgery to include, as appropriate in future meetings, the agencies from which each of you have presently come.

WORKSHOP CHAIRMAN'S CLOSING REMARKS

There are two matters to be taken up before we terminate what appears to me to have been a highly informative series of meetings.

First of all, you will recall that a month or so prior to the Workshop, each of you received a request for your opinions as to the three most important problem areas currently facing personnel selection groups in the Navy. Moreover, you were asked at the same time to pair with each of the problem areas you had indicated as relevant one or more proposed solutions to the problem.

We appreciate the fact that most of the participants provided this opinion data to us in conjunction with transmittal of their itinerary for the meeting. Prior to the onset of the Workshop, your suggested problem areas were categorized by the staff of the Personnel Research Branch of the Laboratory and compiled in a series of Tables, a copy of which has been provided each of the participants. (Appendix A contains a copy of the request for opinion data mailed to each of the participants prior to the Workshop, together with the series of Tables containing the problem areas with their paired solutions, classified according to content.)

Some of you may have wondered why we prevailed upon you to provide us with this opinion material prior to the onset of the Workshop. What we had in mind was to compile a catalogue of relevant problem areas in the selection field (they must be relevant from the standpoint of the participants since they were provided by them), so that in the eventuality that some one or other of the sessions might "run dry," we could then turn to some of the topic areas suggested in the Table to provide a focus for the remainder of the session. Fortunately, I suspect as a function of the comprehensiveness of each session overview as well as the resourcefulness of each participant engaged in the interchanges following the overviews, none of the sessions, according to my observations at least, showed any semblance of losing momentum, much less coming to a halt.

Nevertheless, since the compilation provided by the participants appears to be useful as a means of identifying related topics for future meetings in this broad area, we plan to append the Table to the Workshop Proceedings for your use at a later date. I would like to point out in passing (still referring to the Table of Selection-Related Topics, Appendix A) that in one way or another, most of the classes and subclasses of problems seemed to have come up in one context or another during one or more of the highly productive Workshop Sessions. Though without a doubt, many of these

problems were touched upon all too briefly; nevertheless, regardless of the superficiality of coverage of these many and varied topics, I suspect some very useful "cross fertilization" of ideas probably occurred in all the Workshop Sessions.

Now to the second and final matter. With the question of the feasibility of future meetings in mind, you will recall that a survey form was circulated in the last Session, asking you for your comments, suggestions and criticisms regarding the Workshop in general. We asked you for your opinions as to the "mechanics" of how the Workshop was organized. We asked also at the same time for titles for future meetings of this or other kinds. We are thankful that a number of you responded. Since we have categorized the comments, let me mention some of the comments that seem particularly pertinent for planning future meetings:

Comments Pertaining to the Subject Matter of the Workshop

(1) Six participants were of the opinion that there were too many topics in the present Workshop resulting in too superficial a coverage of each subject matter area,

(2) Too little time was spent in covering in some detail the variety of selection programs represented by the participants,

(3) The topic areas should be more specific and clearly delineated.

Comments Regarding the Mechanics of Organizing the Workshop Sessions

(1) Three people suggested that the overviews would have been more valuable if there would have been prepared papers,

(2) Three people indicated that they felt the overviews were too lengthy and detailed,

(3) Two people mentioned that the technique used by one or two discussion leaders who requested every participant to make a statement regarding the session topic during the interchange led to repetition and to involuntary participation by people whose background was inconsistent with the topic at hand,

(4) One suggestion was to have the discussion leader ask only participants with

detailed knowledge and experience in the subject matter area of concern to take part in the interchange.

Suggestions for Titles for Meetings (Workshops and other Types of Meetings) Related to the Selection Area

- (1) Job analysis of Navy tasks,
- (2) Recent breakthroughs in selection research: Breakthroughs defined as studies providing validity coefficients (cross validated) at satisfactory confidence levels,
- (3) A meeting to clarify the interrelationships between the Naval Activities supporting various kinds of selection programs and the Management Bureaus — BuMed, BuPers, ONR, etc.,
- (4) A meeting to discuss problems involved in the compilation of data banks,
- (5) A meeting to examine the degree to which our research products are being utilized by the "Commanding Officers."
- (6) A conference dealing with development of performance effectiveness criteria and problems involved in obtaining fleet performance data,
- (7) Mathematical-computer models for selection/assignment problems.

One participant only, made a suggestion as to the location of future meetings of this kind, namely to have the West Coast facilities get together for a meeting and similarly, the East Coast facilities do the same thing, the results of each to be pooled.

One final comment had to do with an oversight on our part, namely, our failure to invite a representative from the Naval Examination Center as a participant in this Workshop.

It has been my observation in all of the Workshop Sessions held the past few days that each discussion leader did an outstanding job of, let us say, "setting the tone" for his respective session. Moreover, judging from the frequency and substantive nature of the participant interchanges following them, the overviews also must have served as a strong impetus for participant interaction. Add to this the breadth and experience of each of the participants and the personal enthusiasm each brought to the Workshop Sessions and you have the "blueprint" for a most successful conference.

We greatly appreciate your attendance at these meetings.

NAVAL SUBMARINE MEDICAL CENTER
Workshop on Personnel Selection
26-28 April 1967

PROBLEM AREA — POSSIBLE SOLUTION SHEET

It seems a reasonable assumption that although there are communalities of opinions among selection researchers generally as to **WHAT** constitutes the most (and least) important problem areas in this area of specialization and **HOW** one goes about solving them. Yet, it also seems likely that there are differences.

THEREFORE, we are asking the Workshop attendees to list below: (1) the three most important, (difficult, complex) problem areas in Personnel Selection facing Navy Selection Groups in the next decade, ranking the three from most important (rank 1) to least important (rank 3). Moreover, coinciding with each of these three areas we are asking you to indicate or suggest a possible solution(s) for each of the three areas.

PROBLEM AREAS

POSSIBLE SOLUTION(S)

Most Important

Second Most Important

Third Most Important

ESTIMATED TIME OF ARRIVAL (day/hour)

MODE OF TRAVEL

MOTEL ACCOMMODATIONS REQUIRED Single Double

IF DOUBLE, LIST NAME OF ROOMMATE

Classes & Subclasses	<u>Frequency and Degree of Relevance</u>			Possible Solutions
	Most Important	Second	Third	
<u>2.0 Retention</u>				
2.1 Selecting those who will stay in Navy beyond 1st or 2nd enlistment.	X			Development of screening techniques to identify the poor reenlistment risk.
2.2 Reenlistment		X		Additional three-year obligation for men with costly training.
2.3 Retention of critical personnel.		X		a. Re-examine dependence of military systems on key personnel. b. Study effects of career incentive programs for critical personnel.
2.4 Retention of Officers.			X	Training for a specialty rather than rotating every few years to eliminate disruption of family life.
<u>3.0 Personnel Management</u>				
3.1 Selection and effective utilization of the marginal man.	X			a. Job analysis. b. Relating job analysis to individual differences. c. Increased attention to "moderator" variables in selection research.
3.2 Officer selection from enlisted ranks.	X			Increase motivation to stay in by giving individuals a billet in their specialty.
3.3 Executive selection.		X		Longitudinal studies.

Classes & Subclasses	<u>Frequency and Degree of Relevance</u>			Possible Solutions
	Most Important	Second	Third	
3.4 Improving image of the serviceman in eyes of the public.			X	Improved and expanded public relations program handled by specialists.
3.5 Placement of personnel substandard for a given job into a satisfying job, of value to the Navy.			X	A closer working relationship with BUPERS. Those doing selection research can be more familiar with all opportunities existing in the Navy.
3.6 Utilization of handicapped or substandard personnel.			X	a. Examine effects of special training upon educationally handicapped personnel. b. Predict manpower requirement needs in view of resources.
3.7 Selection of those capable of fitting into advanced technical weapons systems			X	a. Improvement of selection instruments. b. Retention of those who are successful.
3.8 Personnel input to programs.	X X			Maintain high quotas when demand is low i.e., LOW selection ratio.
3.9 Favorable incentives e.g. salary and benefits.			X	Compensation comparable to similarly-trained civilians.
4.0 <u>Validity</u>				
4.1 Increasing in the field validity coefficients of selection information.	X			a. Improved understanding and prediction of behavior. b. Reduction of error variance.

CLASSES OF SELECTION-RELATED PROBLEM AREAS AND SOLUTIONS SUGGESTED BY WORKSHOP PARTICIPANTS

Classes & Subclasses	Frequency and Degree of Relevance			Possible Solutions
	Most Important	Second	Third	
<u>1.0 Criterion Problems</u>				
1.1 Temporal factors related to assessment of psychological fitness and job proficiency	X			Longitudinal system of re-evaluation of personnel.
1.2 Identification of and operational definitions of criteria	X			Need for techniques of job analysis. Use of Q, P, and R techniques of factor analysis.
1.3 Relationship of system performance (e.g. Weapons system) to quality of crew selected.	X			a. Relate weapons or other system performance experience to characteristics of personnel. b. Explore possibility of deterioration of system performance over time.
1.4 Improvement of rating scale data as criteria.		X		c. Use of critical incidents, e.g., accident history. a. Use of factor score estimates derived from rating scale matrices as criterion dimensions rather than ratings themselves.
1.5 Developing relevant multiple criterion indices of performance and adjustment.		X		b. Factor analyze criterion and predictor matrices simultaneously Need for conceptual, theoretical and methodological developments. Necessity of Lab and field research to specify cause-effect relationships and determine their validity and importance.

Classes & Subclasses	Frequency and Degree of Relevance			Possible Solutions
	Most Important	Second	Third	
4.2 Assuring maximum validity for gross screening devices.		X		a. Computerized construction of pencil and paper tests. b. Item bank approach.
4.3 Validation of present selection measures.		X		Follow up studies, Longitudinal studies.
4.4 Increasing predictive validity of clinical interviews.		X		a. Increased clinician awareness of basic requirements of that being predicted through: (1) Job analysis and study of environment. (2) Psychosocial nature of environmental requirements. b. Relating (a) above to individual differences.
<u>5.0 Statistical Models</u>				
5.1 Development of techniques for nonlinear contingency predictors.	X			Use of interaction terms; nonlinear components of relationships; attention to contingency aspects of relationships.
5.2 Need for a model to describe adjustment potential in terms of or as an interactional system of the person with his environment.			X	Use of "hybrid" methods of factor analysis, such as Q-, P-, O-, S- & T-techniques.

Classes & Subclasses	Frequency and Degree of Relevance			Possible Solutions
	Most Important	Second	Third	
5.3 Developing valid multiple predictor model of performance and adjustment criteria.			X	Identification of and operational definition of the most relevant criterion factors.
5.4 Increased attention to role of moderator variables in selection research.		X		Our generally disappointing correlational findings between predictors and criterion may often be a reflection of our failure to recognize that there may be relationships of considerable magnitude in subsamples divided on the basis of some relevant moderator variable(s).
6.0 <u>Application of Systems Analytical Techniques to Selection</u>				
6.1 Development and implementation of comprehensive models for the various personnel systems within the Navy.	X			Discard notion that selection, classification, training, and operations can be treated as separate personnel systems. Assign system responsibilities from procurement to retirement to single labs.
6.2 Improving validities of selection procedures by taking account of environmental and organizational factors in naval assignments.	X			Consideration of specific task and social demands in various naval assignments as they interact with personal needs of crew members and with environmental stresses and supports to determine success of mission and personal costs to individuals.

Classes & Subclasses	<u>Frequency and Degree of Relevance</u>			Possible Solutions
	Most	Second	Third	
	Important			
<u>7.0 Group Processes</u>				
7.1 Crew selection as a concept.		X	X X	a. Crew composition factors in addition to individual factors. b. Mathematical models for describing crew structure.
7.2 Selection of men with maximum potential to adjust within a group during prolonged isolation.		X		One week pre-assignment observation period.
7.3 Development of models for small group structure			X	Model to involve communication nets, information processing, interpersonal attitudes, work-role interdependence, etc.
7.4 Relation of small groups to organizational structure			X	a. Study small naval units in the field. b. Use of small groups computer simulation to derive relevant system properties.
<u>8.0 Miscellaneous Problem Classes</u>				
8.1 Experimentation with new test and measurement technology			X	Automated test booth concept with "on-line" capability.
8.2 Keeping Selection techniques in pace with advancing on-the-job technology	X			a. Consultation with academic world and with industry. b. Use of "brain-storming" sessions or "blue sky ideas" approaches to develop new selection techniques

Classes & Subclasses	Frequency and Degree of Relevance			Possible Solutions
	Most Important	Second	Third	
8.3 Measurement and classification of personality attributes.		X		<p>Personality data and concepts have not been employed with notable success in Navy selection and evaluation programs, probably because of low relevance to general or heavily task-oriented criteria. A generally accepted system of personality constructs is not presently available. Personality measures may have greater relevance for specific adjustment indicators, such as interpersonal compatibility (likability), emotional stability, leadership ability, or the like, than for global performance evaluations, usually heavily weighted with task proficiency components.</p>
8.4 Physiological characteristics as selection indices.		X		<p>Interactional model of physiological indicators with aptitude, demographic and attitudinal measures.</p>
8.5 Improved job analytical techniques		X		<p>Interrelate between-job differences with between-person differences.</p>
8.6 Development of appropriate methodology for analyzing qualitative data.		X		<p>Methods of coding (e.g. sieve and exhaustive codes) qualitative data in terms of the degree of reliability possible to obtain.</p>
8.7 Usefulness of observations from simulated environments.	X		X	<p>a. Use of situational test approaches.</p> <p>b. Laboratory simulation approaches.</p>

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) U.S. Naval Submarine Medical Center, Submarine Medical Research Laboratory		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP	
3. REPORT TITLE PERSONNEL SELECTION IN THE NAVY: Proceedings of a Navy-Wide Workshop, April '67			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Interim report on BuMed Research Work Unit			
5. AUTHOR(S) (First name, middle initial, last name) Benjamin B. Weybrew (ed.)			
6. REPORT DATE 23 February 1968		7a. TOTAL NO. OF PAGES	7b. NO. OF REFS 44
8a. CONTRACT OR GRANT NO.		9a. ORIGINATOR'S REPORT NUMBER(S) SubMedResLab Report No. 511	
b. PROJECT NO. MF022.01.02-9004.01			
c.		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
d.			
10. DISTRIBUTION STATEMENT This document has been approved for public release and sale; its distribution is unlimited.			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY U.S. Naval Submarine Medical Center Box 600, Naval Submarine Base Groton, Connecticut 06340	
13. ABSTRACT This report represents the proceedings of a Navy-Wide Workshop on Personnel Selection in the U.S. Navy. The Workshop took place at the Submarine Medical Center in April of 1967. The thirty participants represented the major Navy activities with on-going personnel selection programs. This report consists of the edited transcriptions of the taped recordings from the seven Workshop Sessions entitled: (1) Brief Updating of Selection Programs by representatives of attending activities; (2) Criterion Development; (3) Personality and Aptitude Measurement in Selection; (4) Biomedical Indices in Selection; (5) Small Groups' Observational Data in Selection; (6) Personnel Allocation Processes; and (7) Selection-related Research Areas and Division of Labor Among the Navy Activities. The format for each Session consists first of an Overview of the specific subject matter area by a pre-selected participant, highly knowledgeable in the field. The Overview is followed by "give-and-take" participant interchanges centering around the session topic. In this context, various selection philosophies are discussed. Too, a variety of selection data, both published and unpublished, are mentioned with appropriate reference information affixed by the Editor. Selection methodologies, statistical techniques, computer strategies, and the pros and cons of various psychometric approaches used in selection are discussed in some one or another of the Workshop Sessions.			

DD FORM 1473 (PAGE 1)

1 NOV 65

UNCLASSIFIED

Security Classification

S/N 0101-807-6801

3ND PPSO 13152

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Personnel selection in the Navy Biomedical indices in selection of personnel						